



Investissement et déréglementation dans le secteur électrique

Benoit Peluchon

► To cite this version:

Benoit Peluchon. Investissement et déréglementation dans le secteur électrique. Economies et finances. Université Panthéon-Sorbonne - Paris I, 2007. Français. NNT: . tel-00226380

HAL Id: tel-00226380

<https://theses.hal.science/tel-00226380>

Submitted on 30 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE PARIS 1 PANTHEON-SORBONNE

ECOLE DOCTORALE D'ECONOMIE PANTHEON-SORBONNE

THESE POUR LE DOCTORAT EN SCIENCES ECONOMIQUES

(Arrêté du 30 mars 1992)

Présentée et soutenue publiquement par

Benoît PELUCHON

INVESTISSEMENT ET DEREGLEMENTATION

DANS LE SECTEUR ELECTRIQUE

Thèse dirigée par Mr le Professeur Roland Lantner (Université Paris 1)

Soutenue le 12 décembre 2007

Rapporteurs : Claude BERTHOMIEU

Jacques PERCEBOIS

Suffragants : Jean-Pierre LAFFARGUE

Roland LANTNER

Jacques LESOURNE

L'UNIVERSITE PARIS 1 PANTHEON-SORBONNE
n'entend donner aucune approbation ni improbation aux
opinions émises dans les thèses ; ces opinions doivent être
considérées comme propres à leurs auteurs

Remerciements

Je tiens particulièrement à remercier mon directeur de thèse, le professeur Roland Lantner, dont le soutien s'est révélé particulièrement précieux tout au long de ces dernières années.

Un certain nombre d'amis ont fortement contribué par leur présence et leur aide à l'achèvement de cette thèse. Qu'ils en soient remerciés, je sais qu'ils se reconnaîtront.

Résumé (français) :

La thèse s'intéresse aux problèmes d'investissement en moyens de production électriques qui sont apparues ces dernières années dans les pays ayant déréglementé cette industrie. Dans la première partie, les différentes crises ayant affecté les marchés de gros sont recensées et analysées, afin de montrer qu'il existe un problème récurrent de sous-investissement en moyens de pointe. Les prix sur les marchés de gros sont tels que les coûts fixes de nouveaux moyens de pointe ne sont pas couverts, phénomène qui a trouvé l'appellation de « revenus manquants » dans la récente littérature théorique. Dans la deuxième partie, un modèle décrivant l'investissement d'un duopole en situation de demande aléatoire est comparé à un modèle de construction similaire décrivant l'investissement d'un monopole public. La comparaison permet de montrer qu'une hausse des prix de l'électricité sur les marchés de gros ne peut palier au problème des « revenus manquants ». Le résultat est étendu au cas d'un n-oligopole. La troisième partie s'intéresse donc aux différents moyens de mettre en place une rémunération complémentaire de la capacité des moyens de pointe. Différents designs de marchés de capacités sont analysés à la lumière des conclusions de la deuxième partie. Enfin, la nécessité de maintenir des capacités comme « réserves opérationnelles » est étudiée. La conclusion, à la lumière d'un article important de Joskow et Tirole, est que celles-ci doivent être considérées comme un bien public. Cet aspect semble empêcher la mise en place d'un marché de capacité, et, ce faisant pose un problème important à la déréglementation du secteur électrique.

Mots-clés (français) : Investissement, Electricité, Modèles Mathématiques

Titre en anglais : Investment and deregulation in the electricity generation industry.

Résumé (anglais) :

This work addresses the issue of investment in the electricity generation industry. As the analysis of many crisis which have affected electricity markets shows, there is a systematic under-investment in peak capacity. Electricity prices are not high enough to cover fixed costs of such generators, a phenomenon that has been dubbed « missing money » in some recent papers (Stoft). The investment decisions of a duopoly facing random demand are then compared to those of a public monopoly. The results are that no prices may be high enough to solve the « missing money » problem, since the duopoly is able to exercise market power in order to maximise his profit. This results systematically in fewer peak capacity in the duopoly case than in the public monopoly case. This remains true in the case of a n-oligopoly. The necessity of designing a mechanism remunerating capacity is thus demonstrated. Capacity markets are then analysed in the light of those results. What appears is that operating reserves are a public good and, as such, prevents capacity markets to solve the « missing money » problem. This casts a shadow on the pursuit of deregulation in the electricity industry.

Mots-clés (français) : Investment, Electricity, Modelization

Centre d'Economie de la Sorbonne

106/112 Boulevard de l'Hôpital

75647 Paris Cedex 13

TEL. +33 (0)1 44 07 81 00

Equipe de recherche MATISSE

SOMMAIRE

Introduction

1- La déréglementation du secteur électrique et l'investissement en moyens de production

- a- Les caractéristiques propres au secteur électrique
- b- Le problème des revenus manquants et les incitations à l'investissement
- c- Les modélisations des marchés électriques

2- Le pouvoir de marché par l'investissement

- a- Choix de modélisation et optimum social
- b- Le duopole déréglementé
- c- Conclusions

3- Les réserves opérationnelles

- a- Le statut particulier des réserves opérationnelles
- b- Les nouveaux designs de marché de capacité
- c- La propriété publique des moyens de pointe

Conclusion

INTRODUCTION

Alors même que le mouvement de déréglementation du secteur électrique, entamé dans les années 90, semble s'imposer dans tous les pays développés, un certain nombre de graves crises ont secoué les marchés électriques ces dernières années et ont semblé mettre en cause leurs modes de fonctionnement. La plus connue étant celle ayant affecté la Californie pendant l'été 2000 (et nécessité une forte intervention fédérale), mais des crises conséquentes ont également affecté l'Allemagne, l'Italie, le Nord-est des Etats-Unis ou bien l'Espagne. Ces crises ont été marquées par la nécessité de « délester » une partie de la demande d'électricité, ou bien par des prix extrêmement élevés dus aux craintes d'avoir à recourir à des délestages.

Comme dans le secteur des télécommunications, la déréglementation du secteur électrique passe par la séparation des activités de production (soumises à la concurrence) des activités de réseaux (qui restent un monopole public avec tarification réglementée). Mais à la différence de celui-ci, l'activité de production du secteur électrique affiche une singularité plus importante, du fait des fortes contraintes techniques qu'imposent la production et la distribution de l'électricité. En premier lieu, l'impossibilité de stocker l'électricité conditionne toute l'organisation de la production, ses coûts et les prix sur les marchés. En effet, l'offre doit être égale en permanence à la demande. Cette dernière étant soumise à de fortes variations, fournir la demande à moindre coût exige un parc de production diversifié et à même de faire face à ces aléas. La déréglementation de la production électrique nécessite donc un ensemble d'institutions ou de procédures particulièrement imposant et qui alimente fortement les débats théoriques en économie.

Or, s'il est une question qui semble se poser avec plus d'acuité depuis plusieurs années, c'est la capacité d'un secteur électrique déréglementé à induire suffisamment d'investissement pour remplir sa mission. Les crises de ces dernières

années ont en effet toutes été provoquées par des déséquilibres entre l'offre et la demande d'électricité. De plus, la déréglementation des secteurs électriques européens et américains s'est effectuée dans un contexte de surcapacité de production. La rationalisation de cet excédent était alors un des arguments en faveur de l'introduction de la concurrence dans la production électrique. Cependant, la capacité des institutions mises en place à favoriser de nouveaux investissements n'a pas fait l'objet de réflexions. C'est précisément le problème qui se pose aujourd'hui et qui a trouvé une reconnaissance académique à travers l'appellation « resource adequacy », ce que l'on peut traduire par sécurité des approvisionnements. Il s'agit de cerner s'il y a bien un problème de sous-investissement dans le secteur électrique, puis de voir quelles en sont les causes et les moyens d'y remédier.

Notre recherche s'inscrit dans une réflexion générale sur le fonctionnement des marchés électriques. De par leur complexité, ces marchés ont généré une importante production théorique faisant appel à plusieurs branches de l'économie (théorie des enchères, économie industrielle, néo-institutionnalisme, etc.), sans pour autant que l'on dispose d'une théorie générale. L'ambition de cette thèse est de proposer des pistes pour construire un cadre théorique permettant une meilleure compréhension du processus d'investissement dans le secteur électrique, et ainsi, de contribuer au débat sur la sécurité des approvisionnements. Il existe en effet très peu d'analyses théoriques du processus d'investissement dans un secteur électrique déréglementé, alors même que cette problématique dans le cadre d'un monopole public a amené de très nombreux travaux, qui ont nourri des débats théoriques plus larges. La grande complexité du mode de fixation du prix de l'électricité dans un secteur concurrentiel complique singulièrement les analyses par rapport au cadre « classique », qui sert encore de référence pour l'évaluation du secteur (mesure des « coûts échoués » par exemple ou réflexions sur le « mix énergétique » en France).

Les modèles développés dans les années 90 pour analyser le fonctionnement des marchés électriques ont été, soit des modèles de concurrence à la Cournot, soit des

modèles dit à « Supply Function Equilibrium » (à fonctions d'offres, les firmes choisissant conjointement prix et quantités). Ces modèles ont pu mettre l'accent de façon intéressante sur la notion de pouvoir de marché (la capacité de certains producteurs à agir sur les prix de façon substantielle et à leur avantage), mais ne se sont pas intéressés au processus d'investissement. Les modèles utilisés ici permettent, tout en respectant l'esprit de ces premiers modèles, de faire un pont entre la théorie des investissements dans un cadre public et le fonctionnement des marchés électriques dans un cadre déréglementé. C'est pourquoi les comportements d'investissement des producteurs vont systématiquement être comparés aux résultats d'une maximisation du bien-être social.

La première partie de la thèse, après un rappel des caractéristiques intrinsèques du secteur électrique et de leurs conséquences économiques, examine ce qu'on a appelé le problème des « revenus manquants » (« missing money »). En effet, plusieurs économistes (Joskow (2006 a et b), Cramton et Stoft (2006)) avancent que les marchés électriques ne rémunèrent pas suffisamment les moyens de pointe. Tous les autres segments de production sont également affectés, dans la mesure où ils dépendent également des revenus dégagés pendant les périodes de pointe. Ce problème peut avoir plusieurs origines, mais la principale pointée par la littérature semble être l'existence de prix plafonds qui réduisent de fait les profits dégagés pendant les périodes de forte demande. Une estimation des « revenus manquants » sur le marché français est proposée à partir des données de Powernext depuis son lancement en 2002 jusqu'au premier semestre 2006. Le marché français est en effet proche de son niveau de capacité d'équilibre, puisque la construction de nouveaux moyens de production a commencé et que le RTE prévoit des tensions entre offre et demande à court terme. Enfin, une présentation des différentes tentatives de modélisation des marchés électriques est effectuée, avec un accent particulier mis sur les concepts proposés par Stoft (2002), qui les a conçus pour essayer de remédier au problème des « revenus manquants ».

La deuxième partie propose un modèle de concurrence imparfaite à la pointe où le niveau des capacités de production est décidé par les entreprises. Ce modèle reprend la plupart des caractéristiques importantes des marchés électriques et s'appuie sur les réflexions de Stoft concernant la fixation des prix de l'électricité. La comparaison entre les résultats du modèle et le programme d'investissement d'un monopole public (maximisant le surplus collectif) permet de mettre en évidence le pouvoir de marché que détiennent les entreprises à travers le processus d'investissement, alors même qu'elles ne peuvent influencer directement sur le niveau des prix. Ce pouvoir de marché, différent des autres formes de pouvoir qui ont été étudiées jusqu'à présent, a de nombreuses conséquences sur les dispositions éventuelles pour remédier aux « revenus manquants ».

La troisième partie, enfin, s'intéresse plus particulièrement aux réserves opérationnelles et à leur importance dans la gestion d'un système électrique. Ces réserves, qui sont des moyens de production choisis par le gestionnaire de réseau suivant certaines conditions, permettent en effet d'éviter un effondrement du réseau électrique en permettant à ce dernier de faire face à des variations de demande en temps réel. De par leur mode de fonctionnement elles peuvent être considérées comme un bien public. La prise en compte de cet aspect complique singulièrement les perspectives de résoudre le problème des « revenus manquants », même en utilisant les nouveaux designs de marchés de capacité explicitement construits pour restaurer les incitations à l'investissement. Le fait que des moyens de pointe soient construits et gérés par le gestionnaire de réseau est alors examiné comme une alternative pour traiter le problème des réserves.

Partie 1 : L'investissement en moyens de production électrique

a- Les caractéristiques propres au secteur électrique

Le secteur électrique présente un certain nombre de caractéristiques qui lui sont propres et qui impactent la problématique de l'investissement. Ces caractéristiques peuvent être rassemblées sous deux principaux items. D'abord, la nécessité de s'intéresser à l'équilibre entre offre et demande d'électricité d'un point de vue physique préalablement à toute analyse économique, du fait de l'importance économique des contraintes physiques. Ensuite, l'importance des aléas qui affectent l'équilibre offre-demande d'un point de vue physique et d'un point de vue économique.

i. Le caractère non stockable de l'électricité

Le trait fondamental duquel découle l'importance de s'intéresser à l'équilibre physique est qu'il est impossible de stocker l'électricité. Par conséquent, à tout instant, la puissance appelée / demandée doit être égale à la puissance produite. La puissance est mesurée en kilowatts (kW), tandis que l'énergie, qui renvoie au produit de la puissance appelée par la durée d'appel, est mesurée en kilowattheures (kWh). Sur une année, par exemple, il est donc nécessaire de disposer d'équipements de production dotés d'une puissance au moins égale à la demande attendue la plus forte. De là découle l'importance de la notion de « système électrique » (ou « plaque de cuivre »), c'est à dire d'un ensemble de moyens de production et de consommateurs suffisamment interconnectés pour qu'il soit possible d'ignorer en première approximation les problèmes de réseaux. On peut en effet à tout instant considérer la somme des puissances des différents moyens de

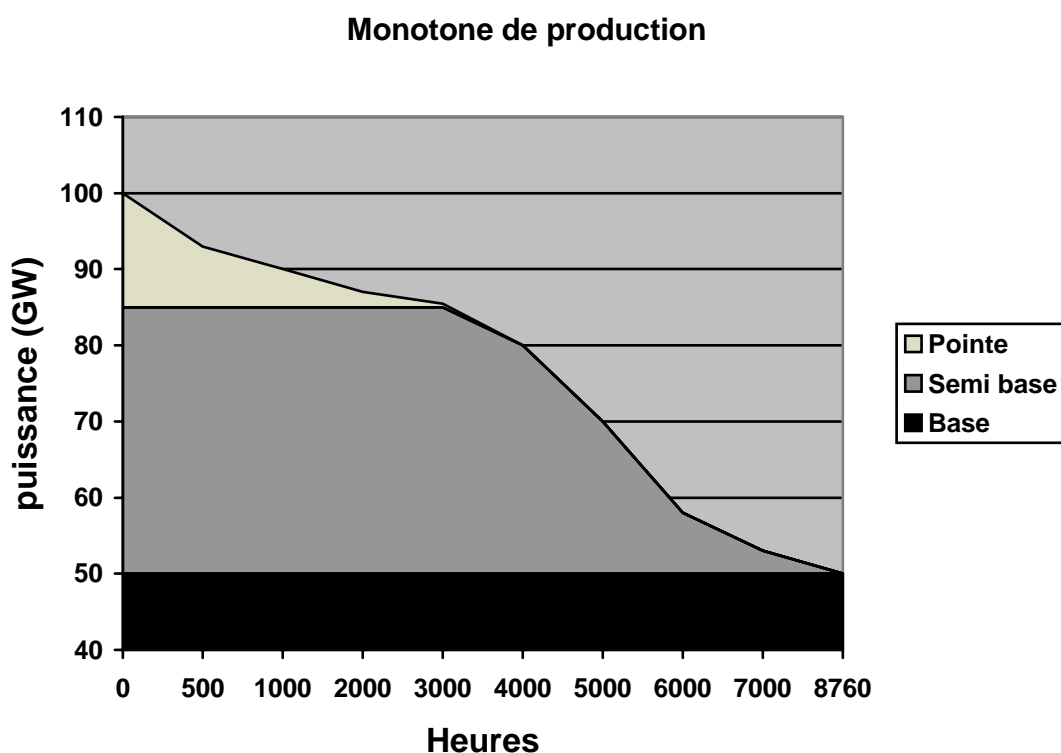
production du système comme représentant son offre, sa demande l'étant par la somme des puissances demandées par les consommateurs. A l'offre définie précédemment, il convient de rajouter la puissance des autres systèmes, mobilisable par le biais des interconnexions avec les systèmes voisins, à la puissance maximale fixée. De même, à la demande il convient de rajouter la puissance demandée par les autres systèmes, là aussi limitée par la puissance des interconnexions éventuelles.

Il s'agit avec cette notion d'ignorer les aspects de localisation dans la détermination de l'équilibre offre-demande. Ainsi, la France (dont on exclut la Corse, puisque non reliée au système français mais au système italien par un câble sous-marin) peut être considérée en première approximation comme un système électrique, qui a des interconnexions avec d'autres systèmes comme la Grande Bretagne, l'Italie, l'Espagne et le système germanique. Ce dernier comprend, quant à lui, l'Allemagne, la Suisse et l'Autriche. Cependant, la notion de système électrique est rendue floue par la difficulté de mesurer les puissances de transit entre deux systèmes A et B qui sont eux-mêmes interconnectés avec un troisième C, tout flux d'électricité de A vers B impliquant des flux de A vers C. Néanmoins cette notion peut être considérée comme opérationnelle. Elle est importante pour définir des zones de prix, mais également des zones de réglementation, puisqu'il n'y a pas à l'heure actuelle de projet européen d'un gestionnaire du réseau européen (comme il en existe un à l'échelle de chaque pays). Toute décision d'investissement dans une zone donnée doit donc être évaluée à l'échelle des fondamentaux du système considéré (qui incluent bien sûr une idée du comportement des autres systèmes aux interconnexions).

Dans un système donné la demande varie beaucoup en intra-journalier (jour/nuit) et suivant les saisons selon les usages prépondérants dans la consommation (chauffage et/ou climatisation et/ou éclairage). Il y a là une source d'importantes différences entre les pays. Ainsi, la France connaît une forte disparité entre la puissance demandée l'hiver et celle demandée l'été du fait de la prépondérance du

chauffage électrique par rapport au chauffage au gaz. Cette disparité est bien moindre dans les autres pays européens, voire inversée en Espagne, où c'est la climatisation qui augmente la puissance demandée l'été. De même, la « pointe » journalière en France se situe vers 19 heures, tandis que la puissance appelée en Allemagne est plus plate dans la journée (avec néanmoins une baisse la nuit). Ces disparités se sont plus ou moins retrouvées dans les tarifs d'électricité dans les pays où la fourniture d'électricité était réglementée. Elles imposent des contraintes de disponibilité des équipements, et surtout ont des conséquences en terme de rentabilité de ceux-ci.

Il est possible, pour une année donnée, d'établir une monotone des puissances, c'est-à-dire une relation entre niveau de puissance demandée et nombre d'heure de demande. Cette monotone se représente aisément sous forme de diagramme (hypothétique) :



Ainsi la puissance demandée 8760 heures pendant l'année est de 50 GW, tandis que la puissance demandée pendant au plus 4000 heures est de 75 GW.

Cette notion de monotone est classique en économie de l'électricité (Varoquaux 1996 ; Laffont 1996). Il faut bien noter que la monotone d'une année future ne peut être établie qu'en espérance.

Avec cette relation il est possible de déterminer les seuils de compétitivité des différents équipements de production. En effet, ces derniers exhibent des répartitions coûts fixes / coûts variables différentes, ce qui leur donne une compétitivité hiérarchisée suivant le nombre d'heures d'appel attendues. Ainsi, une centrale nucléaire comporte de très importants coûts fixes (construction, démantèlement, effets de taille dans les investissements, etc.), mais de faibles coûts variables (uranium et part variable des coûts de la filière combustible). A l'inverse, les centrales au charbon coûtent moins cher à construire, mais plus à exploiter, en raison du coût du charbon. On peut encore citer le moyen de production favorisé à l'heure actuelle qu'est le Cycle Combiné au Gaz (CCG) dont la répartition entre les coûts est encore plus accentuée vers les coûts variables. Les moyens qui ont le coût de production moyen total le plus faible pour une production sur toute l'année (soit pour une année de 365 jours, 8760 heures), sont appelés moyens de production « en base ». Ce sont les moyens qui répondent à la demande correspondant à la base de la monotone des puissances. Pour une production en base, et en tenant compte de l'entretien, le nucléaire est le moyen de production le moins cher pour des prix du gaz élevés, le CCG le devenant pour des prix du gaz médians ou faibles. Le calcul de la compétitivité du nucléaire est néanmoins compliqué par l'importance des économies d'échelles à la construction (effet de filière), la valorisation des coûts de démantèlement et de l'absence d'émissions de gaz à effets de serre. A cela s'ajoute en outre des considérations stratégiques sur la dépendance d'un pays non producteur de combustibles fossiles.

Pour satisfaire le surcroît de puissance, il est nécessaire de pouvoir disposer d'autres moyens de production qui vont fonctionner durant un nombre d'heures

inférieur à 8760. Le coût de production total moyen augmente pour ces fonctionnements du fait des coûts fixes. Il est alors souvent plus économique de faire appel à d'autres types d'équipements à la structure de coût plus variable, tels que les centrales au charbon ou au fuel. Le seuil de rentabilité entre deux équipements se situe au nombre d'heures qui égalise le coût de production total moyen. C'est cet effet qui explique la diversité nécessaire des parcs de production existants. Pour une demande donnée sur une année, le parc correspondant à la monotone des moyens de production est celui qui satisfait la demande et minimise le coût de production de l'électricité (Varoquaux 1996).

Il est dès lors possible de distinguer demande de « base », de « semi-base » et de « pointe ». La demande de base correspond aux niveaux de puissance demandés au moins 7000-8000 heures sur une année, donc de façon quasi-permanente. La semi-base renvoie aux puissances demandées entre 3000 et 5000 heures sur une année, et la pointe aux puissances demandées moins de 3000 heures sur une année. Ce découpage n'est pas officiel, ni clairement défini (sauf pour les produits de marché du type « peak » ou « off-peak » qui incluent alors dans leur définition temporelle les plages horaires précises (8-20 heures par exemple) de certaines journées). Il s'agit d'une convention qui dépend essentiellement des différents moyens de production existants, convention dans laquelle la base correspond à la demande optimisant le fonctionnement des centrales nucléaires (ou CCG), la semi-base aux centrales au fuel, etc. En outre, cette distinction ne présage pas de la répartition des heures d'appel dans l'année, alors qu'il existe des contraintes dynamiques pour la semi-base ou la pointe (« chauffe » des centrales au charbon par exemple).

On appelle enfin réserves l'ensemble des moyens de production électrique dont le « combustible » est un stock fini (mais renouvelable) de coût variable quasi-nul. C'est le cas notamment des centrales hydrauliques qui fonctionnent avec des réservoirs / barrages (par opposition à celles qui se contentent d'exploiter le courant d'un fleuve). Sont également considérés comme des réserves les contrats

interruptibles avec des consommateurs. Certes ces contrats ne sont que des réductions de demande, mais l'essentiel réside dans le fait qu'il s'agit d'un moyen de production dont la gestion est assimilable à celle d'un stock. On a par ce biais une possibilité de stockage de l'électricité, bien que limitée par la puissance des équipements qui utilisent le stock (taille des turbines ou des contrats). Leur gestion s'opère alors sur valeur d'usage puisque, si leur coût variable est généralement nul, l'utilisation d'une ressource limitée a un coût dans le temps, en particulier du fait de la saisonnalité des prix et donc du coût d'opportunité de l'utilisation de ces réserves.

ii. L'importance des aléas

Le deuxième aspect fondamental qui caractérise le secteur électrique est la présence de nombreux aléas physiques et économiques qui affectent l'équilibre entre l'offre et la demande (et l'affectent d'autant plus que l'électricité n'est pas stockable).

La demande

La plus grande incertitude sur le long terme est certainement la demande en énergie. A titre d'exemple, dans les années 70, les prévisions pour la France faisaient état d'une consommation qui avoisinerait les 1000 TWh en 2000, alors qu'elle a été en réalité de moins de 500 TWh. Ce paramètre est un des plus important pour le dimensionnement global d'un parc, et par conséquent pour les décisions d'investissement. Il est particulièrement difficile à prévoir puisqu'il dépend conjointement de la croissance globale de l'économie sous-jacente au système, de l'évolution des usages de l'électricité et du prix de l'électricité à long terme !

Il est essentiel de souligner que la demande d'électricité est très peu élastique au prix à court terme, dans la mesure où elle dépend des besoins d'un certain nombre d'équipements déjà installés pour lesquels il n'existe pas de source d'énergie alternative. Par ailleurs, la consommation des particuliers est dans la majorité des cas tarifiée par les pouvoirs publics, ce qui entraîne une élasticité-prix de leur demande nulle (en dehors de la saisonnalisation éventuelle des tarifs). Quand les particuliers peuvent changer de fournisseurs, le problème reste identique, puisqu'ils ne peuvent connaître les variations en temps réel des prix de l'électricité et adapter leur consommation en conséquence. Il existe néanmoins des consommateurs pour qui l'électricité est un poste budgétaire important et qui peuvent se permettre de réduire temporairement leur consommation. Ces consommateurs sont la source d'une forme d'élasticité de la demande d'électricité contre rémunération (ex : les tarifs Effacement des Jours de Pointe en France). Néanmoins, en dépit de ce cas, une part très importante de la demande n'est pas sensible aux prix. Ainsi, Joskow (2006b) estime que 98 % de la demande d'électricité aux Etats-Unis est inélastique. C'est pourquoi la part élastique de la demande est souvent modélisée comme un moyen de production supplémentaire (tout effacement de la demande à partir d'un niveau de prix fixé est assimilable à un générateur dont le coût proportionnel est égal à ce niveau : les EJP sont modélisés par EDF comme des moyens de pointe gérés sur valeur d'usage comme une retenue d'eau).

Cette quasi-inélasticité de la demande d'électricité a du point de vue économique des répercussions majeures. Elle supprime de fait toute pression concurrentielle de la part des consommateurs sur les producteurs, et cela d'autant plus que l'électricité ne peut être stockée. Cette caractéristique du secteur électrique ouvre donc la possibilité aux producteurs d'exercer un pouvoir de marché, ce qui se traduit par des prix plus élevés. Ce pouvoir est à l'origine de nombreux dispositifs institutionnels visant à l'amoindrir, il occupe une place centrale dans les débats sur la déréglementation du secteur électrique.

Le climat constitue un facteur exerçant une influence notable sur l'équilibre de court terme. En effet, pour un système donné, en fonction de la nature et de la

répartition des usages (chauffage, climatisation, éclairage), la consommation va connaître des aléas plus ou moins marqués. Ainsi, la part importante du chauffage électrique en France (par opposition à une prédominance du gaz dans le reste de l'Europe) entraîne une forte dépendance à la température extérieure de la puissance demandée en hiver. A l'inverse, un pays à la température moyenne élevée, mais aux étés particulièrement chauds, équipé de nombreuses installations de climatisation, connaîtra une demande plus aléatoire en été.

La disponibilité des moyens

Tout comme la demande, l'offre est également soumise à de nombreux aléas, de coût bien sûr, mais aussi de nature physique. L'ajustement en temps réel entre puissance demandée et puissance offerte bute naturellement sur l'écueil des pannes d'équipements. Celles-ci ne sont pas rares dans l'industrie électrique dans la mesure où les équipements sont justement soumis à de nombreuses contraintes physiques dues au processus de production lui-même (combustion, turbinage) qui relèvent de l'industrie lourde. Les normes de sûreté jouent également un rôle non négligeable, car elles peuvent empêcher un équipement de fonctionner, cette problématique se posant avec une acuité particulière dans le cas du nucléaire.

La plupart des équipements nécessitent de nombreux entretiens, aussi bien réguliers qu'exceptionnels (décennal, par exemple, pour des maintiens de permis d'exploitation). Ces entretiens représentent une incapacité à produire qu'il convient de placer au mieux dans l'année afin d'exploiter son parc optimalement. Là encore, ce problème est démultiplié dans le cas du nucléaire en raison de la longueur des arrêts (également plus aléatoires) et de la nécessité de programmer des arrêts pour rechargement en combustible (ce qui n'est pas le cas d'une centrale charbon par exemple). Ainsi, l'entretien annuel d'une centrale au charbon (semi-base) ou au fuel (pointe) est « masqué » pour des centrales qui ne tournent qu'une partie de l'année, puisqu'il a lieu dans les périodes où elles n'ont que très peu de chances de fonctionner.

Enfin, les moyens de production hydrauliques et éoliens sont soumis à l'irrégularité éventuelle de leurs apports en eau ou vent. Par exemple, la récente crise au Brésil (2001) était essentiellement due au fait que la grande majorité de la production électrique brésilienne est fournie par des usines hydroélectriques, lesquelles dépendent de réservoirs dont le remplissage suit un cycle pluri-annuel et qui s'est révélé beaucoup moins important que la moyenne. En Europe, l'énergie hydroélectrique n'occupe une place centrale que dans trois systèmes : la France, le système germanique (du fait de la Suisse) et le système scandinave. Ce dernier est particulièrement dépendant des apports hydrauliques. Quant aux éoliennes, un de leur grand défaut est qu'il n'est pas possible de prévoir leur capacité de production, et ce malgré un effet de foisonnement, en particulier pour les parcs éoliens off-shore, en plein développement en Allemagne. Un parc qui dépend d'éoliennes doit dès lors nécessairement disposer de moyens de substitution en cas d'absence de vent, ce qui revient à cumuler des coûts fixes.

On appelle « marges de réserves » la puissance (proportionnelle à la demande) qui doit être disponible en plus de la demande. Leur rôle est d'éviter une rupture de l'équilibre offre-demande en cas d'aléas négatifs. De plus, elles rendent différents services indispensables à la gestion d'un système électrique (ce qu'on regroupe en général sous l'intitulé « réserves opérationnelles »).

Les conséquences sur l'investissement

La nature particulière de la production d'électricité a de nombreuses conséquences quant à la façon de raisonner en matière d'investissement, en particulier pour les moyens de pointe, qui, au sein d'un système donné, ont à assumer le rôle d'une assurance contre la défaillance physique (les coupures).

Ainsi, l'instabilité des prix des matières premières qui servent de combustibles rend également aléatoire la hiérarchie des différents moyens de production. Ce problème s'avère d'autant plus important qu'il concerne des producteurs en concurrence sur un même marché, puisqu'un type de centrale qui devient moins

compétitive qu'un autre peut voir son fonctionnement fortement réduit, ne trouvant plus sa place dans la monotone. Cet effet existe également du fait du progrès des nouvelles technologies de production. C'est ce qui s'est passé avec l'introduction des Cycles Combinés au Gaz (CCG) au rendement beaucoup plus élevé que les anciennes turbines à combustion, et qui ont rapidement remplacé les centrales au charbon tant que moyen de production de base privilégié en Angleterre. Les revenus d'un investissement en sont rendus plus aléatoires. En conséquence, la diversification des risques conduit les producteurs à favoriser un parc hétérogène afin de capter au mieux les variations des coûts. Ce problème concerne peu les moyens de pointe dans la mesure où ceux-ci correspondent déjà aux moyens de production les plus chers en terme de coût unitaire. Ils ont donc peu de chance de se voir déclassés par une variation des prix des matières premières, mais ils restent concernés par l'introduction de nouvelles technologies plus efficaces.

Une complication importante des investissements dans des moyens situés en haut de la monotone réside dans le fait que les débouchés futurs de ces moyens dépendent des investissements réalisés dans les moyens situés plus bas dans la monotone. Ainsi, un investissement dans un moyen de pointe peut se révéler non rentable si des capacités supplémentaires non anticipées sont introduites en semi-base. Les revenus futurs d'un actif de pointe dépendent donc largement des anticipations des investissements dans les autres secteurs de la monotone. Leur rentabilité prévisionnelle est dès lors extrêmement aléatoire et conditionnée par la bonne appréciation du risque de déséquilibre par manque d'offre.

Enfin, dernier facteur de complication : l'importance des coûts fixes et la politique de prix du producteur. Le coût marginal d'un kilowattheure d'électricité produit par une centrale est en effet égal à son coût proportionnel de production. Il exclut donc les coûts fixes qui sont importants dans le secteur électrique. Dans le cadre d'un secteur public, la tarification se fait souvent au coût marginal rehaussé d'un facteur permettant de couvrir l'intégralité des coûts fixes (tarification de Ramsey-Boiteux par exemple, Boiteux 1956). Au sein d'un marché déréglementé, la situation est plus confuse (Borenstein 2000). Une firme « price taker » est prête à produire dès que le prix sur le marché dépasse son coût marginal. Son profit peut

donc être positif comme négatif, suivant le niveau du prix. Ce dernier, dans une situation où aucune firme n'a de pouvoir de marché, est égal au coût marginal du système de production dans son intégralité. Il est alors égal au coût proportionnel du dernier moyen démarré. Si celui-ci appartient à la semi-base, le prix sera alors supérieur au coût moyen total de production des moyens de base, qui sont rentables. On parle alors de rentes infra-marginales, pour désigner les revenus issus de prix égaux aux coûts marginaux du système électrique, mais supérieurs au coût variable du moyen considéré. Pour une centrale de pointe, le seul moyen de récupérer ses coûts fixes est de vendre à un prix supérieur à son coût proportionnel, ce qui peut être malaisé si de nombreuses capacités sont disponibles. On comprend dès lors l'intérêt pour un producteur de rechercher des situations où il dispose d'un pouvoir de marché, même si l'existence d'un tel pouvoir n'est en rien indispensable théoriquement à une unité de pointe pour couvrir ses coûts fixes. Nous aurons l'occasion de revenir plus en détail sur cet aspect.

D'une manière générale, la fixation des prix sur un marché de l'électricité est donc éminemment complexe et étroitement corrélée à la structure de coût de l'ensemble des producteurs et à leur éventuel pouvoir de marché.

iii. La déréglementation

La déréglementation du secteur électrique est un phénomène relativement récent qui présente un certain nombre de différences nationales. Pour autant, les réformes s'inspirent toutes d'un même cadre théorique qui justifie la mise en concurrence de certains des services auparavant assurés par des monopoles publics. Ce cadre, issu des théories sur les monopoles contestables et les économies de réseaux, postule pour le secteur électrique la séparation entre d'un côté des activités de production, mises en concurrence, et de l'autre des activités de réseau, qui restent monopolistiques et publiques. Les activités de production sont alors organisées autour de marchés de l'énergie électrique, marchés au travers desquels se fixent les prix de gros de référence. Ces marchés constituent de ce fait l'élément central de l'économie de la production électrique déréglementée. On s'intéressera donc au fonctionnement de ces marchés et à leur organisation à travers deux points de vue : celui de leur organisation institutionnelle dans un premier temps ; celui des crises qui les ont affectées ces dernières années dans un second temps.

Les différents types de marchés électriques

Dans un système électrique déréglementé, l'ajustement entre offre et demande s'opère sur des marchés, et plus particulièrement sur des marchés de gros. Cela signifie que l'électricité produite est achetée aux producteurs préalablement à son acheminement aux consommateurs finaux par des sociétés ad hoc. Il s'agit donc bien de marchés de gros puisque les quantités des transactions sont importantes, et que les prix de cession ne reflètent que les coûts de production, sans considération aucune du transport. Le marché fonctionne donc comme si le réseau de transport auquel sont rattachées les unités de production se résumait à un point, l'activité de dispatching indispensable au fonctionnement du marché étant confiée à un tiers unique.

Le nombre de marchés et leur organisation diffèrent d'un pays à l'autre. Néanmoins une organisation a minima se retrouve invariablement. Pour tenir compte des spécificités de la production d'électricité, deux marchés physiques au moins cohabitent : un premier marché dit « spot » ou « veille pour le lendemain » et un marché dit « d'ajustement ».

Le marché spot est celui sur lequel s'effectue l'essentiel des transactions. Il détermine en J-1 les volumes et les prix des transactions qui seront effectuées en J. Ce décalage est nécessaire pour les besoins du dispatching sur le réseau électrique, qui nécessite un certain nombre de calculs et des prévisions des besoins de puissance pour assurer le bon fonctionnement du réseau. Les transactions se font sur tous les pas de temps de la journée (une demi-heure en général). Elles s'effectuent en général (ce n'était pas le cas sur le Pool anglais, Green 1999 et Glachant 1998) comme sur un marché financier classique en réalisant des transactions avec annonces de prix préalables, le prix spot correspondant au pas de temps étant le prix le plus élevé des transactions réalisées.

Le marché spot fonctionne donc sur la base de prévisions de la demande la veille pour le lendemain. Prévisions qui peuvent évidemment être sujettes à erreur, d'où la nécessité du marché d'ajustement (balancing market) qui assure l'équilibre en temps réel entre offre et demande, aussi bien pour les erreurs de prévision que pour l'ajustement au sein d'un pas de temps, opérations qui nécessitent des équipements réagissant très rapidement (turbine hydraulique par exemple). Un troisième type de marché peut également exister pour des transactions qui ont lieu entre la veille pour le lendemain (marché spot) et le temps réel (marché d'ajustement), répondant ainsi au besoin de pouvoir ajuster des positions sans nécessairement intervenir sur le marché d'ajustement (ce qui impose un certain nombre de contraintes).

Enfin, parallèlement aux marchés physiques, existent également des marchés de produits dérivés : futures, forward et options. En outre, du fait de la non stockabilité de l'électricité, prix à terme et prix spot ne sont pas liés aussi

étroitement que sur un marché financier. Ils reflètent donc en partie les anticipations des prix spot futurs.

Au delà du schéma général précédent, les marchés existants présentent un certain nombre de différences significatives d'un pays à un autre. En effet, ils reposent sur un ensemble d'institutions et de normes nationales qui leur permettent de fonctionner. Et comme l'expose Glachant (1998) pour l'organisation du Pool, celles-ci empêchent une transposition aisée du mode d'organisation d'un pays à un autre. Certains pays ont néanmoins fait office de « laboratoire » pour les réformes adoptées par d'autres pays.

Ainsi, le premier marché d'électricité à avoir été organisé sur la base d'un marché spot est le Pool anglais issu de la réforme du système électrique anglais de 1990 et 1991. Il a de fait joué un rôle d'expérimentation d'un certain nombre de comportements et réglementations pour les autres pays qui ont mis en place leurs propres réformes ultérieurement. C'est lui qui a donné lieu dans la littérature économique à la plupart des tentatives de modélisation d'un marché de l'électricité. Pour autant, il a été remplacé en 2001 par le NETA, du fait de manipulations de la part des acteurs anglais du marché, qui fonctionne sur une base plus proche d'un marché de matières premières traditionnel.

Le deuxième modèle important d'organisation est Nordpool, créé en 1993, nom générique d'un ensemble de marchés communs à la Norvège, la Suède, le Danemark et la Finlande. Parallèlement, les régulateurs des pays membres de Nordpool se sont rassemblés dans l'organisation Nordel dont la fonction est d'établir une position commune à ces pays sur les sujets concernant le secteur électrique. Par extension, Nordel désigne le système électrique commun à cet ensemble de pays.

On distingue en général trois grands types de marchés organisés (Finon 2001) : le marché obligatoire, dont l'archétype est le Pool anglais, le marché volontaire (ex : Nordpool) et les bourses électriques.

Le premier type repose sur l'obligation qu'ont les producteurs de vendre toute leur électricité sur le marché, en affichant pour chaque pas de temps, les prix et les quantités auxquels ils sont disposés à effectuer des transactions. La coordination du parc du système concerné s'opère dès lors de façon proche du « merit order » traditionnel (appel des unités dans l'ordre croissant de leurs coûts). Ainsi, sur le Pool il n'y avait même pas confrontation de l'offre et de la demande, la procédure d'appel était réalisée à partir d'une prévision de demande faite par le régulateur. Les prix auxquels avaient été effectuées les transactions n'étaient d'ailleurs connus que quelques jours plus tard du fait de la complexité du mode de calcul de la rémunération des producteurs (incluant le coût marginal du système, les coûts de démarrage et un terme de rémunération de la disponibilité) d'une part, et du décalage entre la prévision de la demande et sa réalisation d'autre part. L'OMEL en Espagne et la Borsa Elettrica italienne fonctionnent encore sur ce modèle, considéré comme peu transparent et sujet à manipulation de la part des producteurs. L'un des points clés du fonctionnement du Pool résidait d'ailleurs dans le fait que la rémunération des producteurs pour une tranche horaire donnée était le prix proposé par la dernière unité effectivement appelée dans le dispatching. C'est pourquoi, il existe en général dans ce type de marchés des règlements ad hoc afin d'éviter que des unités soient déclarées indisponibles sur le marché spot afin d'effectuer des transactions à des prix plus élevés sur le marché d'ajustement. En effet, des manipulations de ce type ont été relevées sur le Pool et ont conduit à une pénalisation de ces comportements en 1991, après investigation de l'Offer (Green, Newbery 1996). C'est d'ailleurs la conviction du régulateur anglais que de telles manipulations continuaient sans pour autant qu'il soit possible de les prouver, qui a conduit au remplacement du Pool par le NETA (New Electricity Trading Arrangement) en mars 2001, sur un modèle plus proche de Nordpool.

Ce modèle fait du marché un complément des transactions effectuées par contrats bilatéraux. Il repose sur une confrontation décentralisée des déclarations d'achat et de ventes par heure. Son mode de fonctionnement se rapproche ainsi beaucoup de celui des marchés de matières premières traditionnels, et permet de donner un référent de prix à l'ensemble des transactions. Ce modèle est souvent considéré comme le plus transparent, d'autant plus si, comme pour Nordpool, ce sont les Gestionnaires du Réseau de Transports qui possèdent le marché et en assurent le bon fonctionnement.

Le troisième type de marché est celui des bourses électriques : ils s'agit de marchés privés complètement décentralisés (transactions dites « over the counter »). Ainsi plusieurs marchés peuvent cohabiter, comme c'est le cas en Allemagne (EEX, LPX). Les informations sur les transactions sont ensuite données au gestionnaire du système qui procède au dispatching journalier. Comme le rappelle Finon (2001), ce système est celui qui pousse le plus loin la rupture avec les contraintes de la régulation technique journalière, et par conséquent, avec le modèle du monopole public centralisé.

Des dysfonctionnements répétés sur les marchés électriques

Plusieurs crises ont affecté les systèmes électriques déréglementés au cours des dernières années, alors même que le mouvement de déréglementation s'accélérait. Certaines de ces crises ont été particulièrement médiatisées (Californie), en raison notamment du caractère spectaculaire des coupures dans la fourniture d'électricité d'un des Etats les plus riches des Etats-Unis. D'autres n'ont pas connu la même couverture médiatique, bien qu'elles fussent peut-être révélatrices de dysfonctionnements plus graves du modèle de déréglementation dans sa généralité. La plupart de ces crises sont issues de difficultés d'ajustement entre offre et demande et méritent donc que l'on revienne sur leur déroulement. Deux crises ont particulièrement retenu notre attention. La première consiste en la forte hausse des prix de gros sur le marché allemand (représentatif de la plaque germanique qui inclut également la Suisse et l'Autriche) en décembre 2001, prix qui guident les transactions en France en l'absence d'un marché suffisamment développé. La seconde est la crise californienne de 2000-2001, marquée par des coupures importantes de la fourniture et une faillite des distributeurs. Le premier exemple est présenté comme une crise bien que des prix élevés lors de période de pointe ne soient pas choquants per se. Mais leur niveau exceptionnel et la possibilité envisagée d'une défaillance (ce qui se traduit par le non respect des marges de réserves) permet d'envisager un dysfonctionnement du marché, légitimant ainsi l'emploi du terme « crise ».

La pointe de décembre 2001 sur la plaque germanique

La « crise » de décembre 2001 sur la plaque germanique (Allemagne, Suisse, Autriche) n'en est pas vraiment une dans la mesure où il n'y a pas eu de coupures

de courant. Les marchés allemands ont donc toujours équilibré offre et demande. Pour autant, il y a eu une forte tension qui s'est traduite par des prix spot extrêmement élevés sur les heures de pointe de plusieurs journées de décembre, jusqu'à 700 €/MWh (le prix du MWh pour une fourniture en base annuelle était de l'ordre de 23 € pour 2001 sur la plaque germanique), les prix spot en base montant à 275 €/MWh le mercredi 19 décembre (100 et 225 les deux jours précédents). De tels niveaux de prix sont élevés par rapport aux coûts marginaux des équipements de pointe (le coût variable d'une TAC (turbine à combustion) peut être de l'ordre de 80 €/MWh avec les prix du pétrole de l'époque). En revanche, ils peuvent se justifier si l'on incorpore dans les prix un amortissement des coûts fixes (35 €/MW/an à peu près pour une TAC pour une durée de vie de 20 ans et un taux d'actualisation de 8 %). Il semble donc que les prix atteints reflètent une stratégie de prix agressive de la part des offreurs, puisqu'ils ne se justifiaient pas au niveau des coûts marginaux. Par ailleurs, les prix à terme pour les années 2002 et 2003 ont également légèrement augmenté (entre 1 et 2 €/MWh), indiquant une transmission de ces problèmes sur le prix de la base. Ce dernier point indique clairement que les niveaux constatés n'étaient pas ceux d'une simple pointe de demande.

Ces prix élevés sur le marché allemand ont également provoqué des hausses sur les marchés spot anglais, jusqu'à 60 €/MWh (l'Angleterre n'est interconnectée qu'à la France), espagnols, jusqu'à 50 €/MWh (l'Espagne est également uniquement interconnectée à la France), et français, bien que les volumes sur ce dernier soient très faibles et non significatifs (les niveaux des prix des transactions en France sont accrochés à ceux des marchés allemands). En fait, il y a même eu des coupures de courant pour les consommateurs en Espagne pour insuffisance d'offre, les tensions conduisant le gestionnaire de réseau espagnol à annuler unilatéralement des exports de courant vers la France afin d'essayer de rétablir l'équilibre. Il semble donc qu'il y ait eu contagion des tensions du système germanique vers le système espagnol à travers le système français. De fait, les causes principales de la crise se situent dans le déséquilibre du système français.

En effet, si le mois de décembre a été très froid et s'est traduit par des consommations record en France et en Espagne et par des niveaux élevés dans le

système germanique, c'est la consommation française qui, de tous ces pays, est la plus sensible aux variations de température. Le niveau beaucoup plus élevé en France que chez ses voisins de l'équipement en chauffage électrique implique une dépendance beaucoup plus forte de la demande d'électricité à la température. En décembre les températures ayant atteint des niveaux significativement plus bas que la normale saisonnière, la France a donc connu une hausse également plus importante que la normale de sa consommation. Par ailleurs, un nombre élevé de centrales nucléaires étaient à l'arrêt, privant le système français d'une puissance élevée et l'obligeant à devenir importateur net d'électricité, quand il est en général exportateur net. Se sont donc conjugués sur les marchés allemands un déficit d'offre conséquent puisque la France n'exportait plus, et même importait massivement, et une demande légèrement plus élevée. C'est également ce qui s'est produit en Espagne, conduisant même à des défaillances de fourniture, sans pour autant que les prix montent aussi haut que sur les marchés allemands, ce qui tendrait à confirmer l'idée d'une stratégie agressive des offreurs sur ces marchés, là où en Espagne les offreurs pouvaient craindre plus de mesures de rétorsions du régulateur (un comportement qui semble exister sur le Pool anglais dans le milieu des années 90, comme le montre empiriquement Wolfram 1999). Car, dans les débats qui ont suivi en Espagne, c'est le sous investissement des producteurs qui a été incriminé face à une croissance de la demande plus forte que prévue. Des tensions entre offre et demande s'étant déjà manifestées auparavant, il semble que cela soit vrai, l'inversion des flux avec la France ne faisant que précipiter la crise.

Le rôle de la France dans les événements ne semble pas avoir été perçu de tous les observateurs, un quotidien français affirmant même que le système français continuait à exporter de l'électricité, tandis que des traders cités par Platt (lettre d'information professionnelle sur les marchés d'électricité) s'interrogeaient sur des comportements spéculatifs, voire des manipulations de cours. En fait, cette crise met en évidence la dépendance des systèmes voisins de la France à ses exports, compétitifs du fait de leur origine nucléaire, mais qui peuvent se transformer en imports dans un cas extrême d'aléas défavorables (température et disponibilité du nucléaire). Or, ce type de crise pourrait se renouveler plus fortement à l'avenir

dans la mesure où la dépendance évoquée semble devoir se renforcer. En effet, le système allemand devrait faire face à des déclassements de puissance, tandis que le système français ne semble pas se lancer dans des investissements nouveaux conséquents et que la demande devrait continuer à augmenter.

La crise californienne

La crise californienne a été abondamment médiatisée pour des raisons diverses et souvent non économiques : Etat mythique, des coupures de courant dans un des Etats le plus riche du monde et à la pointe de la libéralisation, etc. Néanmoins, elle présente également un intérêt indéniable pour l'analyse du fonctionnement d'un système déréglementé, notamment du point de vue du comportement d'investissement de producteurs privés. Si la raison principale des coupures tient plus à la faillite des distributeurs d'électricité induite par une déréglementation mal pensée (les distributeurs devant vendre à un prix fixé par la loi et acheter à un prix de gros complètement libre), on peut constater en amont l'existence d'un déséquilibre entre offre et demande dû conjoncturellement à une vague de chaleur (augmentant la consommation pour climatisation) et à des possibilités d'import réduites (faibles précipitations dans les Etats voisins), et structurellement à une hausse de la demande sur plusieurs années sans que la puissance installée augmente significativement.

En effet, Faruqui, Chao, Niemeyer, Platt et Stahlkopf (2001) notent qu'entre 1993 et 1999 la demande de pointe a augmenté de 18 % par an tandis que la capacité installée augmentait de 0,1 % par an seulement. Entre 1996 et 1999, la demande croît de 5500 MW et l'offre de 672 MW. Joskow (2001) permet de comprendre les raisons de ce sous investissement. La déréglementation du secteur électrique californien a lieu en 1998, mais le début des réflexions et des débats publics y conduisant ont débuté en 1993. Une des raisons qui a poussé à s'engager dans une réforme du système électrique, est le niveau jugé trop élevé des marges de réserve

(capacité utilisable, mais inemployée pour faire face à un aléa) à l'époque. Pendant la période 1993-97, aucune construction de nouvelles capacités n'a été engagée, du fait du flou entourant la future réglementation du secteur. Quand le cadre a été enfin connu, une course pour l'obtention des autorisations de construction a eu lieu entre investisseurs, mais elle s'est heurtée à un délai d'obtention très important en Californie, lié entre autres à des contraintes environnementales. Ainsi, la première mise en service n'a pu avoir lieu que pendant l'été 2001, soit trop tard par rapport à la hausse de la demande rendue critique par les aléas de température et d'hydraulicité.

Trebing (2001) met également l'accent sur les manipulations de la part des producteurs et des traders qui ont délibérément fait monter les prix en augmentant artificiellement les prix du gaz, dont certains contrôlaient l'approvisionnement. L'affaire Enron a mis la lumière sur les stratégies de goulot d'étranglement en certaines portions du réseau détenues par Enron ainsi que les stratégies de manipulations des prix plus classiques. Dans un article de 1999, Borenstein et Bushnell concluaient déjà à l'existence d'un pouvoir de marché de la part des producteurs en Californie, sur la base d'une modélisation du marché de gros californien utilisant des données historiques sur les coûts des producteurs. La situation de sous investissement, bien qu'apparemment non déclenchée par le comportement des producteurs, a mis en place les conditions de leur pouvoir de marché. C'est pourquoi les prix ont pu atteindre des niveaux si élevés sur le marché de gros en 2000 et 2001, jusqu'à 175 USD par MWh comme prix spot sur la journée en août 2001, soit des niveaux de prix comparables à ceux de la crise de décembre 2001 sur la plaque germanique.

Vers de nouvelles crises ?

D'autres crises importantes se sont également produites depuis 2000 : des hausses de prix importantes dans le nord-est des USA (notamment à New York) ou en

Angleterre dans le mois qui a suivi la mise en place de NETA ; l'apparition de pics de prix durant l'été en France et certains de ses voisins. Ainsi, durant l'été 2003, la canicule qu'a connue l'Europe s'est traduite en France par des prix très élevés (1000 €/MWh le 11 août pour le bloc 10-16 h sur Powernext), ainsi que sur le marché spot d'Amsterdam (2000 €/MWh sur certains blocs). Ces prix sont issus d'un déficit en puissance du parc de production français, confronté à une demande beaucoup plus élevée que la normale pour un mois d'août, du fait de l'utilisation de climatiseurs. Autrement dit, la France a été obligée d'importer de l'électricité des systèmes voisins, à n'importe quel prix, compte tenu des niveaux atteints. Or, bien qu'anormalement élevée, la puissance demandée est restée bien en deçà de ses niveaux hivernaux. Il est possible d'expliquer ce phénomène en soulignant l'importance de la part du nucléaire dans le parc de production français. Chaque tranche d'une centrale nucléaire nécessite d'être arrêtée périodiquement pour être rechargée en uranium ou être inspectée pour des raisons de sécurité. Ces arrêts, parfaitement prévisibles, au contraire des avaries pouvant survenir, font l'objet d'une planification afin de pouvoir avoir lieu aux moments les plus opportuns, c'est-à-dire en été. Il est donc plausible qu'un certain nombre d'arrêts aient été programmés cet été, immobilisant une partie importante de la puissance installée. Ces pics ne sont donc pas suffisants pour parler d'un manque de capacités installées, ce que semblent confirmer les niveaux de prix en 2004. Or, l'année 2005 et le premier semestre de l'année 2006 ont à nouveau été marqués par de fortes turbulences. En 2005, dans la première moitié du mois de mars (300 €/MWh pour l'heure 17, le 07 mars) et sur la fin de l'année (600 €/MWh pour l'heure 19 le 29 novembre) ; en 2006 à nouveau pendant un épisode caniculaire durant l'été (à nouveau 1000 €/MWh pour l'heure 12 le 26 juillet, et 416 €/MWh pour le bloc « peak », ce qui est exceptionnel, étant donné que ce bloc « peak », qui recouvre les heures , correspond à 3000 heures par an, ce qui relève plus de la semi-base que de la pointe à proprement parler). Pour pouvoir comparer ces niveaux de prix aux précédents, il faut néanmoins garder à l'esprit que les prix du pétrole et du gaz ont connu de très fortes augmentations en 2005 par rapport aux années antérieures, et donc que les coûts variables des unités de pointe n'ont pas du tout les mêmes

ordres de grandeur (avec un baril de pétrole à 60 \$, le coût variable d'une TAC au fioul se situe vers les 200 €/MWh). Néanmoins, les niveaux atteints et leur occurrence de plus en plus régulière sont des signes assez fiables d'une tension croissante entre offre et demande en France.

Les crises qui ont été mises en avant ne sont donc pas des événements isolés. Plus largement, la question se pose de savoir si dans un système déréglementé les producteurs sont vraiment incités à investir en moyens de pointe.

Car, si la déréglementation semble ne pas poser de problèmes d'investissement significatifs pour répondre à la demande « standard » (base et semi-base au sens large) au vu de l'expérience de la déréglementation anglaise (de nombreux CCG ont été construits ; entre 1990 et 1996, on recense 11,2 GW de nouvelles capacités pour un marché de l'ordre de 60 GW (Green et Newberry 1996)), les problèmes de pointe de demande et de prix semblent prendre de l'importance avec la déréglementation, puisque de tels événements n'avaient lieu que très exceptionnellement dans le cadre d'un secteur électrique public centralisé. Et en effet, les hausses de prix importantes lors de tensions peuvent être beaucoup plus rentables que le surcroît de ventes qu'auraient peut-être généré des équipements supplémentaires évitant ces tensions. C'est ce que semble indiquer Joskow (2001) dans son analyse de la crise californienne où il souligne que les marchés de gros marchent mal quand l'excès d'offre est limité. C'est ce point précis que l'on va essayer de mettre en exergue dans la partie suivante.

b-Le problème des « revenus manquants » et les incitations à l'investissement

Plusieurs articles récents (Cramton et Stoft 2006, Joskow 2006b) s'accordent sur le constat d'une insuffisance des investissements dans des systèmes électriques déréglementés. Cela proviendrait, toujours selon les mêmes articles, d'un problème d'incitations à l'investissement, problème qui a commencé à avoir une reconnaissance théorique sous le nom de problème dit de « missing money », que nous avons choisi de traduire par « revenus manquants » (tout en laissant les guillemets à chaque fois que ce terme est utilisé). Il semble, en effet, que les prix sur les marchés de gros n'atteignent pas des niveaux suffisamment élevés lors des pics de prix. Or, de tels niveaux sont indispensables pour couvrir les coûts fixes des unités de pointe, mais constituent également une partie non négligeable des revenus des unités de base et de semi-base. La cause la plus souvent évoquée est l'introduction des prix plafonds sur les marchés de gros, provenant de la volonté de lutter contre la possibilité d'exercice d'un pouvoir de marché. Le problème des « revenus manquants » semble à l'heure actuelle constituer l'explication la plus plausible aux crises récentes.

Une place particulière doit être ici réservée aux travaux de Stoft, dans la mesure où il est le premier à avoir mis l'accent sur l'importance des pics de prix, l'importance d'avoir une politique de prix afin de « contrôler » les paramètres de ces derniers, et enfin, l'importance d'intégrer ces éléments au design des marchés électriques. Cette partie doit donc être l'occasion de revenir sur le mode de fixation des prix de gros dans un premier temps, puis d'essayer de cerner quelles sont les causes du problème des « revenus manquants », pour enfin tenter de mettre en évidence l'existence de ce problème sur le marché français.

i. Le problème des « revenus manquants »

Sur un marché de gros électrique fonctionnant de façon concurrentielle (donc, sans pouvoir de marché exercé), les prix doivent être égaux aux coût marginaux. Comme souligné précédemment, les coûts marginaux sont égaux aux coûts proportionnels des derniers moyens appelés, ce qui, étant donné l'importance des coûts fixes propres au secteur électrique, a souvent conduit à une mauvaise compréhension des mécanismes permettant aux producteurs de couvrir leurs coûts fixes (Borrenstein 2000, Stoft 2002). En effet, l'égalité prix / coût marginal est suffisante pour assurer cette couverture, sauf pour les moyens de production situés en extrême pointe. Le coût marginal intègre en effet une partie des coûts fixes des moyens de production antérieurs à la dernière unité appelée dans le « merit order ». On parle alors de rente infra-marginale. Celle-ci est égale à l'écart entre le coût marginal du système considéré et le coût proportionnel d'un tel moyen de production. Cette rente infra-marginale est suffisante pour couvrir les coûts fixes des unités de base ou semi-base. Le cas des unités de pointe est en revanche plus problématique. En effet, supposons qu'il n'existe que trois types de centrales de production : une de base, une de semi-base, une de pointe. L'égalité prix / coût marginal conduit bien à l'existence d'une rente infra-marginale pour les deux premiers types de centrales, mais pas pour la dernière. Celle-ci doit pouvoir vendre son énergie à un prix supérieur à son coût proportionnel (et donc au coût marginal du système), si elle veut couvrir ses coûts fixes (on parle alors de « rente de rareté » (scarcity rent) pour désigner les revenus dégagés par de tels prix). Une telle situation est évidemment assez fréquente à la pointe, puisque le recours aux dernières unités de production disponibles est le signe d'une tension certaine entre offre et demande, ce qui se traduit toujours par des prix plus élevés que les coûts proportionnels. D'une part, en raison de la politique de prix des producteurs de pointe ; d'autre part, à cause du coût induit par une rupture éventuelle de l'équilibre offre-demande pour tout le système. Dès lors, les incitations à payer

d'un certain nombre de demandeurs peuvent atteindre des niveaux assez élevés, ce qui se reflète dans les prix. Néanmoins, de telles situations de tension ont une occurrence très aléatoire. De plus, plus une unité se situe en haut du « merit order », moins elle a de chances d'être appelée, plus la période lui permettant de couvrir ses coûts fixes est limitée.

Ceci met en lumière l'importance des pics de prix pour qu'il y ait des incitations suffisantes à investir en moyens de pointe (et, dans une moindre mesure dans les autres moyens de production). L'introduction de prix plafonds a donc pu avoir des incidences importantes sur les conditions d'investissement, en limitant la rémunération potentielle des unités de pointe.

Ce qu'on appelle problème des « revenus manquants » est défini ainsi par Cramton et Stoft (2006) :

« The missing money problem is not that the market pays too little, but that it pays too little when we have the required level of reliability. »

Autrement dit, même si un système électrique présente un niveau de capacité proche de ce qui est considéré comme optimal, les revenus générés par les pics de prix sont insuffisants pour couvrir les coûts fixes des unités de pointe. Ce constat est partagé par Joskow (2006a), dans un article important par l'ensemble des faits qu'il recense :

« Moreover, while capacity obligations and associated capacity prices that are components of the market designs in the Northeast produce additional net revenue for generators over and above what they get from selling energy and ancillary services, the existing capacity pricing mechanisms do not appear to yield revenues that fill the “net revenue” gap. That is wholesale prices have been too low even when supplies are tight. »

Joskow essaye de quantifier le revenu manquant à partir de données sur le PJM (incluant les revenus tirés des services ancillaires) et conclut que sur une période allant de 1999 à 2004, seulement 40 % des coûts fixes annuels auraient pu être couverts pour une nouvelle unité de pointe d'une durée de vie de 20 ans. En intégrant les revenus des différents mécanismes destinés à rémunérer les capacités sa conclusion est inchangée. De plus, ce constat vaut pour les autres marchés aux Etats-Unis :

« Every organized market in the US exhibits a similar gap between revenues produced by energy markets and the fixed costs of investing in new capacity measured over several years time. There is still a significant gap when capacity payments are included. »

La seule exception qu'il relève est le marché de New York, tout en notant que les coûts d'une nouvelle installation sur ce marché sont peut-être sous estimés. Ce constat est maintenant reconnu au niveau de la FERC (Cramton et Stoft 2005, Joskow 2006a), ce qui la conduit à recommander systématiquement d'inclure des mécanismes de rémunération des capacités dans les designs de marchés électriques. Il semble en revanche inquiétant de constater qu'aucune recommandation en ce sens n'ait été faite par un régulateur européen.

Les causes généralement avancées de cette situation sont à rechercher du côté des prix plafonds et de leurs conséquences. Ce fait commence à être reconnu chez les régulateurs américains. Ainsi, le rapport au CAISO (California Independent System Operator) de Hogan mentionne clairement les prix plafonds comme origine d'un manque de revenus pour un investisseur. Il en est de même dans le livre blanc d'août 2005 de la CPUC (California Public Utilities Commission) consacré aux marchés de capacités, qui revient largement sur les propositions de Cramton et

Stoft pour un nouveau design de marché de capacité essayant de traiter explicitement le problème des « revenus manquants ».

Les prix plafonds ont été introduits pour lutter contre le pouvoir de marché que pourraient détenir certains opérateurs sur les marchés de gros, qui font artificiellement monter les prix lors des situations de tension. Cela est d'autant plus facile à la pointe, lorsqu'un producteur se retrouve en situation de monopole pour la fourniture des derniers mégawatheures. Mettre en place un prix plafond permet donc de supprimer l'intérêt qu'il pourrait y avoir à retirer de la capacité disponible pour faire monter les prix (ceux-ci étant alors très sensibles à la moindre variation des capacités disponibles sur le marché). Un certain nombre de preuves de la réalité de telles pratiques ont été exhibées lors du procès d'Enron ou sur le Pool anglais (qui existait jusqu'en 2001), ce dernier incorporant des paiements de capacités calculés à partir de la probabilité de défaillance multipliée par le coût de défaillance (VOLL). Il était alors très facile pour un producteur de déclarer défaillantes certaines unités, et de faire fortement augmenter ses revenus (obtenus pour toutes ses unités) quand les marges de réserves devenaient faibles.

Cependant, il est très difficile de dire si un niveau élevé de prix résulte de l'exercice d'un pouvoir de marché ou bien de la recherche légitime d'une rente de rareté en vue de couvrir ses coûts fixes. En effet, les pics de prix sont très aléatoires, et des niveaux de prix qui peuvent se justifier lors d'une pointe, si celle-ci est la seule de l'année, ne peuvent plus l'être si de tels niveaux sont à nouveau constatés lors de pointes ultérieures. Un prix plafond permet donc d'éviter des manipulations qui seraient particulièrement difficiles à déceler, mais au risque de déprimer les rentes de rareté qui constituent une incitation indispensable aux investissements.

Remarquons enfin qu'il est nécessaire d'instaurer un prix plafond afin de remédier aux défaillances du marché (Stoft 2002). En effet, l'inélasticité et le caractère fortement aléatoire de la demande, couplés au fait que les capacités sont limitées, font que l'offre et la demande d'électricité peuvent ne pas s'égaliser. Il existe donc

des circonstances dans lesquelles un marché électrique ne peut parvenir à un prix d'équilibre.

Cependant, Joskow (2006b) note que, lorsqu'on étudie le niveau des prix de gros, les prix plafond sont rarement contraignants et qu'ils ne sont donc pas la seule cause des « revenus manquants ». Il remarque que, pendant la plupart des heures où un déficit de réserves opérationnelles est constaté (« scarcity hours »), les prix n'atteignent pas le niveau plafond. Selon lui, les raisons du problème des « revenus manquants » sont plus à chercher dans les imperfections de marché qui existent pendant ces mêmes heures. Ainsi, lorsqu'il existe un déficit de réserves, l'offre et la demande sont verticales et confondues (puisque'il est nécessaire de réduire les réserves pour éviter des délestages, tant que c'est possible). Dès lors, les prix sont extrêmement sensibles à toute action du gestionnaire de réseau. Joskow indique que, pour la quasi-totalité des heures où existe un déficit de réserves, la demande se situe à un niveau qui n'impose pas de délestages, mais oblige à réduire les réserves. Or, dans ce cas, la dernière mesure prise avant délestages consiste à réduire de 5 % le voltage du système. Cela permet de réduire la demande, mais se traduit également par une baisse des prix, alors même que la situation impliquerait des prix plus élevés. Le coût social de la réduction du voltage (qui implique une dégradation du fonctionnement des équipements électriques) n'est pas inclus dans les prix. Une autre imperfection qui tend à déprimer les prix sur les marchés de gros, est le recours par les gestionnaires de réseau à des achats hors-marché. En effet, le gestionnaire de réseau a fréquemment besoin de pouvoir utiliser certains moyens de production dont les caractéristiques intrinsèques (vitesse de réaction, localisation, par exemple) empêchent de passer par le marché. Ces transactions entre le gestionnaire et certains producteurs ne sont donc pas incluses dans le niveau des prix de gros, alors même qu'elles se font à des prix élevés (du fait de leur nécessité). Les contraintes de gestion d'un système électrique empêchent donc les prix d'atteindre un niveau suffisamment élevé durant les périodes de tension. C'est pourquoi Stoft recommande, entre autres mesures, d'élever les prix de gros

au niveau du prix plafond dès qu'un déficit de réserves opérationnelles est constaté.

Le problème des « revenus manquants » semble donc provenir d'un ensemble d'imperfections des marchés électriques lorsque ces derniers connaissent des tensions offre-demande (c'est-à-dire dès qu'il existe un déficit de réserves opérationnelles). Le prix plafond, nécessaire pour réduire la possibilité d'exercice d'un pouvoir de marché pendant ces périodes, relève donc des outils de gestion des ces imperfections au même titre que les transactions hors-marché. Mais, dans la mesure où il n'est pas la cause principale des « revenus manquants », il devient plus difficile d'espérer corriger ce problème par le biais de mécanismes de marché.

ii. Mise en évidence du problème sur le marché français

A partir des prix relevés sur le marché « day-ahead » de Powernext, il est possible de se faire une idée de l'ampleur du problème dans le cas de la France. Elle connaît une situation très particulière comparativement aux autres pays européens : l'ancien monopole public n'a pas été démantelé et détient une position très prédominante à même de lui permettre d'exercer un fort pouvoir de marché. Or, cette situation rend l'exercice encore plus intéressant dans la mesure où, s'il démontre qu'il y a bien un manque de revenus pour les centrales de pointe, ce phénomène existe alors même que le principal producteur détient un pouvoir de marché suffisant pour lui permettre de faire monter les prix (ce qui a tendance à rendre les investissements possibles).

De plus, le marché français, bien qu'historiquement excédentaire en énergie, semble confronté de plus en plus à des tensions entre offre et demande lors d'épisodes climatiques « tendus ». Ainsi, l'année 2005 a vu des niveaux de prix plus élevés que les années précédentes, tendance confirmée et amplifiée sur la première moitié de l'année 2006. De plus, ces épisodes se sont traduits par des importations nettes d'électricité de la part du système français, alors même qu'il est excédentaire le reste du temps. On peut donc penser que le niveau de capacité optimal pour le système France n'est pas loin d'être atteint, comme aurait tendance à le prouver la mise en route de nouveaux projets d'investissement (un cycle combiné au gaz est actuellement en voie de construction), ainsi que les recommandations du RTE dans son rapport sur la PPI (Programmation Pluriannuelle des Investissements). Ce dernier préconisait de nouveaux investissements dès 2007. Le fait que le système français soit excédentaire en énergie (et exporte) en dehors des périodes de pointe relèverait donc plus d'un surdimensionnement de la base (nucléaire, ce qui rend les exports d'autant plus intéressants pour les niveaux de prix du gaz et du pétrole actuels, ces deux

combustibles donnant le prix de la base dans les systèmes voisins) par rapport au reste du parc de production français.

Il faut donc évaluer le nombre d'heures durant lesquelles aurait pu fonctionner une centrale de pointe et pour quels revenus. Le premier paramètre est relativement simple à évaluer : il suffit de recenser le nombre d'heures pour lesquelles le prix spot est supérieur au coût variable d'une unité de pointe (TAC au gaz ou au fioul, selon les travaux de la DGEMP dans le rapport « Coûts de référence de la production électrique » de décembre 2003, qui sert de référence pour les calculs nécessaires aux décisions de planification énergétique). La deuxième grandeur nécessite de supposer que l'ajout d'une unité de pointe supplémentaire n'aurait pas modifié les niveaux de prix, ce qui est une hypothèse assez forte étant donné l'extrême sensibilité des prix lors des périodes de pointe. Néanmoins, il est extrêmement difficile de faire autrement, elle a donc été retenue. De plus, cette hypothèse tend à sous estimer le problème des « revenus manquants » dans la mesure où l'ajout hypothétiques de capacités de pointe ne peut se traduire que par une baisse des prix. La rente de rareté va donc être évaluée par la différence entre le prix spot et le coût variable d'une unité de pointe. Le problème principal est de connaître le montant des coûts fixes d'une unité de pointe. Le rapport sur les coûts de référence donne différents niveaux de coûts fixes suivant les taux d'actualisation pour une TAC au fioul mise en service en 2007, d'une durée de vie de 15 ans (les TAC au gaz étant pour l'heure absente du parc de production français). Pour un taux d'actualisation de 8 %, ce qui est en général retenu pour les travaux du type PPI, mais ne correspond pas au taux de rentabilité exigé par un investisseur privé, les coûts fixes sont de 328 €/kW. Pour un taux d'actualisation de 11 %, plus en phase avec les exigences d'un investisseur dans un secteur électrique déréglementé, ils sont de 333 €/kW. On peut donc retenir un niveau de 330 €/kW, ce qui donne des coûts fixes annuels de 22 €/kW par an (soit 22000 €/MW par an). Par ailleurs, ce rapport n'a retenu que des hypothèses assez basses pour les prix du gaz et du pétrole (l'hypothèse la plus haute retenue pour le prix du baril de pétrole est de 31 \$ le baril), alors que ces prix ont depuis connu de fortes

hausse (des niveaux de 70 \$ ont été atteints en 2006). On a donc retenu plusieurs niveaux de coûts variables afin de pouvoir effectuer des comparaisons.

Suivant le niveau de coût variable retenu, on obtient les résultats suivants :

Coût variable 100 €/MWh	Revenus annuels pour une TAC (€/MW)	Heures de fonctionnement
2006	17518	410
2005	14866	349
2004	0	1
2003	12415	85
2002	36	7

Les revenus annuels pour 2004 sont en effet de 0,004 € pour une heure de fonctionnement potentielle, et donc plus proches de zéro.

Coût variable 125 €/MWh	Revenus annuels pour une TAC (€/MW)	Heures de fonctionnement
2006	11432	108
2005	9332	145
2004	0	0
2003	10581	67
2002	0	0

Coût variable 150 €/MWh	Revenus annuels pour une TAC (€/MW)	Heures de fonctionnement
2006	9377	61
2005	6354	95
2004	0	0
2003	9076	54
2002	0	0

Il est d'ores et déjà possible de noter la tendance à l'augmentation du nombre d'heures de fonctionnement potentiel en 2005 et 2006, alors même que ces deux dernières années ont connu des mois d'hiver relativement « chauds » (du fait de l'importance du chauffage électrique en France, la pointe de demande est particulièrement sensible aux variations de température). L'année 2006 n'est, bien entendu, représentée que pour les huit premiers mois de l'année (l'été a été inclus du fait des pics de prix exceptionnels qu'il a connu), ce qui semble rendre d'autant plus remarquable le fait qu'elle présente plus d'heures de fonctionnement potentielles que 2005 pour un coût variable de 100 €. Cependant, le premier semestre 2006 a été marquée par une forte augmentation du prix du pétrole, ce qui se traduit mécaniquement par des coûts marginaux du système France plus élevés. Et en effet, pour un coût variable plus élevé, le nombre d'heures de fonctionnement potentiel sur le premier semestre 2006 est inférieur à celui de 2005, tout en laissant présager un haut niveau pour l'année 2006 dans son intégralité.

Les chiffres de la colonne « revenus annuels pour une TAC » sont à comparer aux coûts fixes d'une TAC exprimés en €/MW par an, dont on a vu qu'ils se chiffraient à 22 000 €/MW par an. Aucune des années étudiées, quel que soit le niveau de coût variable retenu, ne permet d'atteindre ce niveau. Les huit premiers mois de l'année 2006 laissent envisager cette possibilité, si le coût variable est de 100 €/MWh, mais comme on l'a vu précédemment, cette année a justement connu des niveaux de prix du baril de pétrole incompatibles avec cette hypothèse. De plus, même en extrapolant la tendance constatée sur les huit premiers mois de 2006, les revenus dégagés par une unité de pointe seraient insuffisants pour couvrir les coûts fixes de celle-ci, alors même que de nouveaux équipements sont en construction.

On peut constater empiriquement à quel point la rente de rareté est aléatoire, certaines années se traduisant par des revenus nuls pour un nouvel actif de pointe. Si les trois premières années étudiées avaient servi pour l'élaboration du business plan d'un investisseur (avec toutes les réserves que doit présenter le raisonnement sur une période aussi courte), la décision d'investissement aurait sans nul doute été négative. Or, les années suivantes présentent des signes de tension offre-demande évidente. Il apparaît fondamental d'intégrer la dimension du risque lié à la construction de nouvelles capacités dans une analyse du problème des incitations à l'investissement. Le marché français semble donc bien connaître une situation tout à fait caractéristique d'un problème de « revenus manquants », alors même que le niveau de capacités installées n'est pas loin de l'optimum.

Un tel constat appelle plusieurs remarques : tout d'abord, il est plus que probable que les hauts niveaux de prix atteints lors des différents pics précédemment étudiés soient de plus en plus fréquents, alors même que la déréglementation du marché français n'est pas achevée, ce qui augure d'un risque politique non négligeable quant aux phases finales de celle-ci. D'autant plus que le pouvoir de marché dont dispose EDF (qui dispose de plus de 90 % du parc de production français) dans un marché sous tension est particulièrement important. Ensuite, il semble nécessaire d'engager une réflexion sur les moyens de mettre en œuvre une politique essayant

de remédier au problème des « revenus manquants ». A cet égard, le dispositif non contraignant de la PPI semble inadapté. Enfin, il est particulièrement urgent de mettre en place de telles réponses, compte tenu du temps nécessaire à la construction de nouvelles capacités. Il est donc fort probable que le marché français de l'électricité connaisse de fortes turbulences dans un avenir proche.

iii. Les problèmes de risque

Un autre problème important impactant les décisions d'investissement en pointe, est le problème du risque lié aux rentes de rareté. De par la position d'un actif de pointe en haut du « merit order », la demande de pointe est une demande résiduelle. Elle cumule donc les risques affectant tous les échelons inférieurs. Tout surinvestissement à un échelon inférieur limite donc la demande adressée aux centrales de pointe (et symétriquement toute avaries à ces mêmes échelons l'augmente). Par ailleurs, comme on l'a déjà évoqué, les pointes de demande sont fortement aléatoires. L'investissement en moyens de pointe est de ce fait affecté d'un risque supérieur à toute autre forme d'investissement. Dans la perspective d'un investisseur privé, un tel risque n'a d'intérêt que s'il se traduit par une prime de risque conséquente, c'est-à-dire par une espérance de gains supérieure. Or, le raisonnement qui a conduit à reconnaître le problème des « revenus manquants » n'a été conduit qu'en espérance. On a donc un facteur supplémentaire tendant à réduire les incitations à investir à la pointe. Ce risque, d'origine technique puisque intrinsèque aux caractéristiques « naturelles » des marchés électriques, est aggravé par les risques liés aux changements de politique de régulation. En effet, on a pu observer des variations importantes des niveaux des prix plafonds sur les marchés aux Etats-Unis du fait des pressions politiques visant à contrôler les prix élevés atteints lors des différentes crises. Ainsi, pour reprendre l'exemple de la Californie, entre 1998 et août 2001, la FERC a autorisé des prix plafonds de, successivement, 750 \$, 500 \$, puis 250 \$. Durant l'été 2000, le niveau du prix plafond sur le marché de New York est passé de 10000 \$ à 1000 \$, sur des considérations tout à fait justifiées relatives à de possibles déséquilibres avec le PJM (dont le prix plafond était, lui, fixé à 1000 \$). De telles variations, par leur impact sur les rentes de rareté potentielles, sont particulièrement dommageables si l'on souhaite inciter des investisseurs à construire de nouveaux moyens de production. En effet, quel investisseur est susceptible de prendre le risque de voir un business plan remis en question par une intervention imprévue d'un régulateur ? Il y a donc un débat important à mener autour de la fixation du niveau d'un prix plafond, afin de

pourvoir réduire ce risque de régulation en établissant les bases d'une politique de long terme des prix plafonds, politique qui serait peu sujette à révision, exactement de la même façon qu'en macroéconomie il est peu optimal de modifier certains critères politiques de façon intempestive si l'on souhaite que le gouvernement dispose d'une crédibilité suffisante.

Stoft (2002) explique quant à lui que la durée et la hauteur des pics de prix peuvent être fixées par le régulateur. La hauteur résulte bien évidemment de l'existence d'un prix plafond ; la durée, quant à elle, peut être déterminée par la politique de réserves opérationnelles du régulateur. Les réserves opérationnelles sont les capacités non utilisées considérées comme nécessaires pour pouvoir faire face à une contingence affectant le système électrique, aussi bien en terme de maintien de tension qu'en terme de couverture de la demande. Dans un système non déréglementé, la politique des réserves opérationnelles (exprimées en général en un pourcentage des capacités installées au-dessous duquel il ne faut pas descendre) était une considération technique et se traduisait par la mobilisation systématique des moyens nécessaires, quel qu'en soit le coût. Dans un système déréglementé, où la responsabilité des réserves opérationnelles incombe au gestionnaire de réseau qui ne dispose pas de moyens de production en propre, il est évidemment peu optimal de raisonner de la sorte. Mobiliser des moyens pour maintenir les réserves représente un coût pour le gestionnaire, qui doit donc essayer de minimiser ce coût au regard de ses contraintes. On peut considérer qu'il se fixe donc un prix plafond au-delà duquel il renoncera à monter. L'originalité du propos de Stoft réside dans le fait de remarquer qu'en pratique ce prix plafond agit en pratique également comme un plafond sur tous les marchés, par arbitrage. De ce fait, la politique de réserves opérationnelles doit être intégrée aux réflexions sur le prix plafond affectant le marché spot, puisqu'il ne peut y avoir deux prix plafonds.

Stoft part donc du raisonnement selon lequel plus le niveau de réserves opérationnelles exigé sera important, plus les moments où le prix plafond sera atteint seront fréquents. Donc, une politique qui combine un prix plafond relativement bas avec un niveau de réserves opérationnelles élevé se traduit par des pics de prix fréquents et / ou d'une certaine durée, mais d'une amplitude limitée

par le prix plafond. Une telle politique a plusieurs avantages : le prix plafond bas permet de limiter les possibilités d'exercice d'un pouvoir de marché, tandis que la fréquence plus importante des pics de prix réduit le risque lié aux revenus formant la rente de rareté. Le risque lié à la construction d'une centrale de pointe est ainsi restreint, tout en maintenant son espérance de revenus. Restent à déterminer les paramètres de cette politique (niveau du prix plafond et niveau des réserves opérationnelles) qui permettra d'induire le niveau de capacité optimal.

Il existe de facto une forme de régulation des pics de prix par la conjonction des prix plafonds et des achats du gestionnaire de réseau pour maintenir les réserves opérationnelles. Les pics de prix étant les situations où les marchés fonctionnent moins bien du fait des caractéristiques de la demande ; de la même façon que les mégawatheures vendus pendant les pics de prix ont un rôle plus important puisqu'il permettent d'éviter un déséquilibre préjudiciable à tous les agents économiques, quelles que soient leurs préférences. D'une certaine manière, les derniers mégawatheures qui permettent d'éviter cette rupture remplissent le rôle d'une assurance pour le système dans son intégralité et justifient par là même une intervention du régulateur.

Quel est le niveau optimal de capacités ?

Le niveau de capacités optimal dans le cadre d'un monopole public a été théorisé par Boiteux et appliqué dans le cadre de la planification des investissements menée à EDF (Varoquaux 1996). La variable clé qui permet de définir ce niveau est le coût de la défaillance (Value of Lost Load en anglais, ou VOLL). Ce dernier donne la valeur au-delà de laquelle les utilisateurs d'électricité préfèrent être délestés que continuer à consommer de l'électricité. Sa valeur est particulièrement difficile à évaluer, puisque dépendant des préférences des différents utilisateurs. En France, elle est fixée autour des 9000 €/MWh. A titre de comparaison, Stoft (2002) rapporte que sa valeur pour le PJM est de l'ordre de 15000 \$/MWh. Le niveau de capacité optimal est donc atteint quand le nombre d'heures de défaillance en

espérance (valorisées au coût de défaillance) assure la couverture des coûts fixes d'une unité de pointe. La formule donnant cette équivalence est :

$$(C_D - C_V) \times H_{\text{def}} = CF$$

où C_D est le coût de défaillance, C_V le coût variable de la centrale de pointe, H_{def} le nombre d'heures de défaillance et CF les coûts fixes annuels de la centrale, exprimés en €/MW/an.

Il est donc indispensable de connaître la monotone de production pour pouvoir définir le niveau optimal de capacités. Le raisonnement est un raisonnement marginal : il s'agit de trouver le niveau de défaillance tel que le coût d'installation d'un mégawatt supplémentaire de capacité soit égal au coût social induit par la défaillance qu'il permet d'éviter. Le raisonnement est réalisé à partir d'une unité de pointe, puisque c'est à celle-ci qu'incombe le rôle d'éviter la défaillance, mais il ne présume en rien de la répartition qui sera effectuée entre les différents moyens de production. La répartition entre ces derniers se faisant à partir d'une minimisation du coût total de production une fois connu le niveau de capacité nécessaire. Plus le coût de défaillance est élevé, plus le niveau de capacité optimal sera élevé, puisque les délestages évités par un mégawatt supplémentaire de pointe sont d'autant plus coûteux. La formule permet également de se donner un critère physique du niveau de défaillance acceptable dans la mesure où le coût de la défaillance, les coûts d'une unité de pointe et la connaissance de la monotone de production en espérance définissent conjointement l'espérance de durée de défaillance. Ce qui se traduit par exemple par une heure de défaillance tous les dix ans.

Cette définition du niveau optimal de capacités est également opératoire dans le cadre d'un secteur électrique déréglementé. En effet, le coût de défaillance est toujours défini, de même que le niveau de capacité optimal, mais ne sert plus de variable de bouclage du programme d'investissement du monopole, puisque les

investissements sont réalisés par des entreprises privées raisonnant à partir d'anticipations de prix. Il faut donc pouvoir lier ce niveau optimal de capacités aux processus de décision des investisseurs. On voit donc l'intérêt d'une politique de régulation des pics de prix à travers un prix plafond et une politique de réserves opérationnelles, dans la mesure où l'on peut essayer de définir les flux de revenus nécessaires pour inciter les investisseurs à atteindre ce niveau. Comme nous l'avons précédemment souligné, intégrer la dimension de risque lié aux flux de revenus couvrant les coûts fixes des unités de pointe conduit à ne pas fixer le prix plafond au niveau du coût de défaillance, puisque cela se traduirait par une prime de risque trop élevée. Le coût de défaillance est par contre toujours utile pour fixer le niveau optimal de capacités. Il est donc nécessaire de fixer un prix plafond plus bas et d'assurer une durée et une occurrence suffisamment régulière des pics de prix à travers la gestion des réserves opérationnelles. Stoft (2002) propose donc d'estimer la fonction de profit de court terme d'un producteur afin de pouvoir fixer les paramètres nécessaires de telle sorte qu'ils induisent le niveau de capacité optimal. Cette fonction de profit dépend du niveau de capacités installées et permet donc d'obtenir une première approche de la dynamique entre profit et capacités installées. La partie suivante détaillera les modalités précises permettant d'obtenir cette fonction de profit.

La réflexion sur le problème des « revenus manquants » engendré par les imperfections des marchés électriques rend donc nécessaire une régulation des pics de prix. Cette régulation commence à peine à être intégrée aux réflexions sur les designs des marchés électriques, comme le montre le Capacity Market White Paper de la CPUC (2005) ou les récents travaux de Stoft et Cramton (2006) autour du design du marché de capacités de New York. Néanmoins, la reconnaissance grandissante du problème des « revenus manquants » rend cette régulation urgente, comme le montre l'occurrence de plus en plus fréquente des crises sur les différents marchés. Plusieurs questions restent ouvertes : la régulation des pics de prix suffit-elle à rétablir les incitations à l'investissement ou bien faut-il instaurer des marchés de capacités en complément de celle-ci ? Comment les réserves opérationnelles peuvent-elles concrètement contribuer à restaurer les incitations à

investir ? Si des marchés de capacités doivent être instaurés, quel doit être leur design et comment s'articulent-ils avec la régulation des pics de prix ? Il est nécessaire de bâtir un cadre analytique permettant de répondre à ces questions, la partie suivante fait le point sur les différentes tentatives de modéliser les marchés électriques.

c-Les modélisations des marchés électriques

Dans un système déréglementé, une partie importante des transactions est réalisée sur un marché de gros, dont les prix servent de référence pour tout le reste du secteur. C'est donc à travers la compréhension du fonctionnement de ces marchés que peuvent être appréhendés les principaux déterminants de l'investissement. De nombreuses modélisations en ont été tentées, faisant systématiquement appel à des théories issues de la concurrence imparfaite. Un des problèmes de ces tentatives est qu'elles semblent pouvoir s'adapter à toutes les configurations de marché (marché obligatoire comme volontaire), sans tenir compte explicitement du contexte institutionnel, autrement qu'à travers le choix de telle ou telle forme de concurrence imparfaite. Cependant, l'ensemble des déterminants de la fixation des prix sur un marché de gros est beaucoup trop conséquent pour être intégré dans un modèle simple, dont la visée n'est très souvent que de répondre à une problématique elle même réduite. En particulier, la plupart des modèles des marchés de gros ont été conçus pour s'intéresser à l'existence éventuelle d'un pouvoir de marché (et donc, pour savoir comment le déceler) et à ses conséquences en terme de bon fonctionnement des marchés. Ils correspondent bien à la période de mise en place de la déréglementation, où l'enjeu principal est de mettre en place une structure de marché concurrentielle à partir de l'ancien monopole public dans lequel était regroupés tous les actifs de production (ainsi, en Angleterre l'ancien monopole public a été divisé en plusieurs sociétés concurrentes au début des années 90), et cela dans un contexte de sur-capacités et de rationalisation du parc de production. Ces considérations expliquent le peu d'attention porté dans ces modèles aux problèmes d'investissement.

Or, les crises survenues dans la décennie suivante, ainsi que la publication d'un certain nombre de documents des régulateurs anticipant des déficits de capacités dans un futur proche, ont mis au centre du débat la capacité des systèmes électriques déréglementés à inciter à de nouveaux investissements. Ces

contributions peuvent être rassemblées sous l'intitulé « sécurité d'approvisionnement », ce qui correspond aux différents intitulés anglo-saxons recensés par Joskow (2006b) (« supply security », « reliability » ou « resource adequacy »). La littérature théorique traitant de ces questions est encore réduite, mais recèle quelques tentatives intéressantes, en particulier ce qu'on peut appeler « approche en terme de fonction de profit » de Stoft qui permet de simplifier le processus de détermination des prix tout en autorisant un pont avec le corpus théorique « classique » de l'économie de l'électricité, c'est-à-dire pré-déréglementation. Une place particulière lui est donc faite dans les analyses qui suivent.

i. Les modèles de marchés de gros et la problématique du pouvoir de marché

La littérature théorique sur les marchés d'électricité s'est évidemment développée dans les années 90, avec une accélération à partir de 1995, marquée notamment par la multiplication des tentatives de modélisation explicite du marché de la production. Etant donnée l'antériorité du Pool anglais, c'est ce dernier qui a donné lieu à la majorité des modèles et a orienté les débats autour de la notion de pouvoir de marché. Comme souligné précédemment, des soupçons persistants de manipulations à la hausse des prix ont marqué son histoire. C'est donc la possibilité pour les entreprises d'avoir un tel pouvoir que cherchent à démontrer ou invalider un certain nombre d'articles. Les pionniers en la matière sont Green et Newberry qui, dans leur article de 1992, affirment que, du fait du choix du régulateur anglais de répartir les unités de production autres que nucléaires entre deux entités en concurrence, il existe un fort taux de markup du coût marginal dans les prix sur le Pool. Des conclusions similaires sont tirées d'une modélisation en terme d'enchères par von der Fehr et Harbord en 1993. La plupart des articles ultérieurs vont ensuite revenir sur cette question, opérant des raffinements de la problématique initiale sur le Pool (Green 1996 et 1999, Wolfram 1999) et son extension à d'autres marchés, comme la Californie (Bushnell et Borenstein 1999, Bushnell, Borenstein et Knittel 1999) ou la Scandinavie (Andersson et Bergman 1995).

Le modèle de Green et Newberry (1992) reprend le concept de « Supply Function Equilibrium » (SFE) dans un oligopole développé par Klemperer et Meyer (1989) en l'appliquant au cas du Pool. Dans ce cadre, les agents / entreprises n'ont plus pour stratégie une quantité ou un prix, mais des fonctions d'offre reliant les deux. L'avantage d'une telle conceptualisation est de mieux représenter le comportement

d'une entreprise face à l'incertitude d'une stratégie en prix ou quantité, qui ne peut s'adapter aux différents aléas. Ce que montrent Klemperer et Meyer, c'est qu'une incertitude sur la demande dans un oligopole permet de réduire considérablement l'ensemble des fonctions d'offre possibles à l'équilibre (important dans un cadre certain), voire d'en assurer l'unicité. Green et Newberry reprennent l'idée des fonctions d'offre, mais assortie d'une demande non aléatoire et variant dans le temps. La justification en est le mode de fonctionnement du Pool de l'époque, où le dispatching du lendemain est obtenu à partir d'une prévision de la demande et des annonces des producteurs des prix et quantités qu'ils sont prêts à fournir pour chacune de leurs unités. Green (1996, 1999) dans des articles postérieurs, réinterroge les fondements de ce modèle, en argumentant que le concept de SFE permet de tester différents niveaux de compétition sur un marché et d'intégrer les équilibres de Cournot et ceux de Bertrand comme cas particuliers de cette approche. Il utilise notamment cette approche pour étudier le marché des contrats sur le Pool. Pour autant, le problème d'une approche avec SFE est que les équilibres sont souvent multiples. Ainsi, Green et Newberry (1992) n'obtiennent que des résultats qualitatifs sur l'existence d'un pouvoir de marché des deux acteurs principaux du Pool, sans pour autant être capable de mesurer l'étendue de ce pouvoir, et donc s'il y a sujet à une intervention des pouvoirs publics. Le même défaut se retrouve dans Green (1999).

Dans leur article de 1993, von der Fehr et Harbord ont bâti un modèle du marché spot comme un mécanisme d'enchères. Ils en dérivent des conclusions assez similaires à Green et Newberry (1992), mais leur modèle n'exhibe pas d'équilibre. Le problème des approches avec SFE et enchères pour étudier des problèmes d'investissement est qu'ils ne permettent pas d'obtenir une prévision des prix de gros d'équilibre sur lesquels effectuer des anticipations de rentabilité. Elles sont donc de peu d'utilité pour l'analyse des problèmes d'investissement de pointe.

Enfin, la dernière approche utilisée pour modéliser un marché d'électricité est la concurrence à la Cournot. Cette dernière, bien que ne répondant pas à une stricte analogie avec le mode institutionnel des marchés de gros (comme peuvent le prétendre les SFE) présente plusieurs avantages. Ainsi, une concurrence en

quantité est plus réaliste sur un marché d'électricité, puisque le coût marginal de production est croissant avec la quantité et qu'une concurrence en prix ne donne pas d'équilibre sous ces conditions (Pignol 1999). De même, Bushnell et Borenstein (1999) justifient l'utilisation de l'équilibre de Cournot par deux arguments : d'abord l'inertie (« stickiness ») à court terme des plannings de production face à un aléa, faisant en sorte que ce sont plutôt les prix qui s'ajustent aux quantités que l'inverse, mais aussi les fréquentes contraintes de capacités du fait des indisponibilités pour entretien (ou autres) et des investissements en nouvelles capacités. Bushnell, Borenstein et Knittel (1999) utilisent également une concurrence à la Cournot afin de montrer que les indices de concentration sont impropres à déceler un pouvoir de marché dans le secteur électrique. Selon eux, comparativement aux SFE, la concurrence à la Cournot est plus maniable, donne plus souvent un seul équilibre et surestime le pouvoir de marché des entreprises. Ils indiquent également dans une note (sans spécifier d'articles) que les modèles basés sur les équilibres de Nash n'ont souvent pas d'équilibres, et que s'ils en ont un, il est proche de l'équilibre de Cournot.

D'une manière générale, les articles s'attachant à la problématique du pouvoir de marché montrent qu'il existe effectivement des situations où les producteurs sont à même d'orienter les prix de gros. Cependant, ces situations sont soit exceptionnelles (Wolfram 1999), soit évitables par une réglementation adéquate qui n'annule pas le problème mais l'atténue à un coût moindre qu'une « re-réglementation » (Bushnell, Borenstein, Knittel 1999 ; Green et Newberry 1992 et 1996, Joskow). Ainsi, les problèmes de manipulations des prix sur le Pool, dont Green et Newberry pensent en 1992 qu'ils seraient évitables avec un plus grand nombre de producteurs, provoqueront une refonte complète de l'organisation du marché spot et du dispatching anglais. Cette argumentation est presque toujours fondée sur une modélisation à partir de données historiques qui constate que les prix n'ont pas été si éloignés des coûts, ou sur un raisonnement « hors modèle » qui n'est pas théorisé.

Le problème de ces différentes modélisations est de ne pas intégrer la problématique de l'investissement au profit d'une réflexion sur le pouvoir de marché et la structure permettant de traiter ce problème le plus efficacement. Or, par rapport aux préoccupations au fondement de la déréglementation du secteur électrique, la problématique de la sécurité d'approvisionnement a pris une importance croissante au regard des différentes crises qu'ont connus les marchés d'électricité. Certains articles récents ont commencé à intégrer plus explicitement l'investissement dans des modèles du marché électrique.

Ainsi, Crampes et Creti (2005) s'intéressent aux possibilités d'une distorsion des résultats d'une enchère à prix uniforme sur un marché électrique par le biais d'un retrait de capacité de la part de producteurs asymétriques. La demande est inélastique, mais les producteurs peuvent influencer sur les résultats de l'enchère par la restriction de leurs capacités de production, ce qui a plus de chance d'arriver si la demande est déterministe. Dans le cas où cette demande est aléatoire, le risque qu'il y ait des retraits volontaires est faible pour peu que la probabilité d'une demande élevée soit suffisante. L'investissement est intégré par le biais d'un jeu en deux étapes, la première consistant dans le choix de la capacité de production de la part des producteurs. Cependant, cet article ne s'intéresse pas aux conditions d'investissement, mais plutôt à l'impact de ce dernier sur l'enchère permettant de déterminer le prix sur le marché spot. Les deux articles suivants, en revanche, adressent directement les problèmes de sécurité d'approvisionnement et de « revenus manquants ».

Joskow et Tirole (2006) essayent d'analyser les conséquences d'un certain nombre de mécanismes non-marchands introduits sur les marchés électriques, et vus en général comme indispensables à leur bon fonctionnement (comme un prix plafond ou les critères de réserves opérationnelles). Ils construisent donc un modèle de marché électrique « idéal », c'est-à-dire dépourvu des mécanismes précédents, où une partie des consommateurs est à même de réagir en temps réel aux prix sur le marché spot, et calculent quel serait l'optimum de Ramsey dans ce cadre (incluant

les décisions d'investissement). Ils se servent de cet optimum comme un point de référence à partir duquel étudier les liens entre prix plafond et paiements de capacité nécessaires pour restaurer les revenus manquants. Ainsi, ils montrent qu'il n'est pas possible de restaurer l'optimum avec un prix plafond et des paiements de capacité s'il existe un pouvoir de marché sur un segment du marché (à la pointe par exemple) et qu'il y a plus de deux états de la nature ; ou s'il n'existe pas de pouvoir de marché, mais qu'il y a plus de trois états de la nature. Ils étudient également comment la nécessité d'avoir des réserves opérationnelles empêche la réalisation de l'optimum de second rang. Ils en concluent que les réserves opérationnelles sont un bien public, un résultat important sur lequel nous allons revenir en détail dans la troisième partie.

Meunier et Finon (2006), quant à eux, partent du constat qu'il existe un fort risque de sous-investissement de la part des producteurs afin de pouvoir augmenter leurs profits, stratégie qu'ils distinguent de celle consistant à retreindre sa production pour faire monter les prix. Cette stratégie est d'autant plus profitable du fait des caractéristiques des marchés électriques. Ils avancent également que les unités de pointe peuvent avoir le statut d'un bien public, car les incitations à investir n'intègrent pas les bénéfices sociaux d'une surcapacité à la pointe. Ils étudient donc comment la possibilité d'investir de la part du gestionnaire de réseau permet de contrer ce sous-investissement potentiel. Leur analyse est menée dans le cadre d'un jeu en trois étapes : les firmes décident de leur capacités de production, puis le gestionnaire de réseau fait de même, enfin les niveaux de production sont définis. Cette « menace » du régulateur permet d'atteindre le niveau d'investissement optimal. Cependant, dans le cas où la demande est inélastique et que les producteurs sont peu nombreux, il n'est pas possible d'atteindre l'optimum social. Les firmes peuvent en effet réduire leur production si le gestionnaire a trop investi, réalisant ainsi un surprofit.

ii. L'approche en terme de fonctions de profit

Cette approche a été proposée par Stoft dans le cadre de ses travaux sur les incitations à l'investissement. Elle présente un certain intérêt dans la mesure où elle donne les bases pour une modélisation de ces dernières et un ensemble de critères pour une politique innovante de régulation des marchés électrique.

La fonction de profit de court terme et ses propriétés

Il s'agit de construire une fonction donnant l'espérance des profits de court terme d'une unité de pointe à partir des capacités installées et des paramètres politiques tels que le prix plafond et le niveau des réserves opérationnelles. Dans cette approche, la demande est considérée comme complètement inélastique. On considère une unité de pointe pour simplifier le raisonnement, mais celui-ci pourrait tout à fait être mené pour d'autres types de moyens de production. Par ailleurs, raisonner à partir d'une unité de pointe permet de faire un lien entre capacités et rente de rareté similaire au raisonnement qui prévalait dans le cadre théorique développé pour le monopole public.

Cette approche repose essentiellement sur le fait que lorsque la demande dépasse un certain niveau (c'est-à-dire quand elle oblige à restaurer les réserves opérationnelles), le prix va être égal au prix plafond du marché. Dès lors, il est possible d'estimer les revenus dégagés par les pics de prix en fonction de différents niveaux de capacités. La connaissance de la monotone de production (en espérance) est fondamentale pour permettre de construire la fonction de profit. La durée pendant laquelle la demande sera supérieure aux capacités auxquelles on a retranché le niveau exigé des réserves opérationnelles est, par définition, la durée pendant laquelle le prix va être égal au prix plafond. Le produit de cette durée par la différence entre prix plafond et coût variable de l'unité donne les revenus

dégagés par les pics de prix et donc constitue la fonction recherchée. Algébriquement, cela revient à écrire l'équation suivante :

$$CT_{\pi}(K) = D(K - RO^E) \times (P_{cap} - CV) \times K$$

où $CT_{\pi}(\cdot)$ est la fonction de profit de court terme, $D(\cdot)$ est la fonction donnant la durée pendant laquelle la demande sera supérieure à l'argument de la fonction (à partir de la connaissance de la monotone), RO^E est le niveau exigé des réserves opérationnelles, K les capacités installées, P_{cap} le prix plafond et enfin, CV le coût variable de l'unité de pointe. Cette fonction donne des valeurs exprimées en €/MWh.

Cette formule rend possible l'étude des effets d'une politique fixant le prix plafond à la valeur du coût de défaillance sans mettre en place de gestion des réserves opérationnelles sur achat au prix plafond (ce que Stoft appelle « VOLL pricing »). Il suffit de fixer le paramètre RO^E à zéro. Il est donc possible de comparer les recommandations de Stoft quant à une gestion des rentes de rareté et la politique qui prévaut actuellement sur les marchés électriques.

Une des hypothèses fortes de cette approche est de ne pas considérer les revenus éventuels hors pics de prix (c'est-à-dire quand la demande n'excède pas le seuil de déclenchement des achats pour restaurer les réserves opérationnelles, et donc que le prix n'est pas égal au prix plafond). Il s'agit donc d'une simplification importante, qui se justifie par les considérations en terme de risque militant pour un prix plafond bas (qui limite de fait les possibilités que le prix ne soit ni au niveau du coût variable, ni au niveau du prix plafond) et par l'hypothèse implicite qu'un prix intermédiaire ne peut résulter que de l'exercice d'un pouvoir de marché, ce qui est discutable, compte tenu des contraintes qui pèsent sur une unité de pointe pour couvrir ses coûts fixes.

Le niveau d'équilibre des capacités est donné par l'intersection de la fonction avec les coûts fixes de l'unité de production exprimés en €/MW/an, puisqu'à ce niveau, l'unité de pointe réalise un profit économique (le taux de profit étant pris en compte par le biais du taux d'actualisation intégré dans le calcul des coûts fixes). Pour un niveau de capacité inférieur, elle réalise un « sur-profit » ; pour un niveau de capacité supérieur, un profit qui ne couvre pas ses coûts fixes. Il est ainsi possible de régler les paramètres politiques (P_{cap} , RO^E) de telle sorte que le niveau d'équilibre de la fonction de profit corresponde au niveau optimal des capacités installées. Dans le cas où le prix plafond est égal au coût de défaillance et le niveau de réserves opérationnelles égal à zéro, l'égalité est réalisée par définition. Cette approche offre donc la possibilité d'étendre le raisonnement développé pour un monopole public au cas d'un marché déréglementé, par le biais d'un raisonnement sur le prix plafond.

Elle permet également une première approche de la dynamique des investissements, dans la mesure où le raisonnement opéré est identique à celui d'un investisseur cherchant à arbitrer entre plusieurs opportunités d'investissement. Celui-ci n'investira que s'il anticipe une rentabilité supérieure à son coût du capital (inclus dans le calcul des coûts fixes), c'est-à-dire si son anticipation du niveau des capacités installées lui assure d'être du bon « côté » de la fonction de profit. Evidemment, la différence principale entre le raisonnement qui conduit à la fonction de profit et celui d'un investisseur est l'incertitude affectant le niveau des capacités totales (donc celui du comportement d'autres investisseurs éventuels). Cependant, l'approche en terme de fonction de profit suppose implicitement dans sa vision de la dynamique la réalisation d'investissements dès qu'il y a « surprofit ». Or, les délais de construction d'une nouvelle centrale de production empêchent évidemment une réaction immédiate des investisseurs, de même que cette approche ne tient pas compte d'un éventuel comportement stratégique des producteurs déjà en place qui peuvent volontairement sous-investir afin de maintenir les revenus générés par leurs unités existantes.

Un des intérêts de construire la fonction de profit de court terme d'une unité de pointe est de pouvoir avoir une approche du risque lié à l'investissement dans une telle unité. En effet, la fonction de profit nous donne les informations sur la façon dont évoluent les profits autour du niveau optimal de capacité. Dans la mesure où la monotone utilisée pour construire cette fonction n'est que l'espérance de monotone, et étant donné les forts aléas entourant la demande, les profits effectifs d'un investisseur peuvent être insuffisants pendant plusieurs années, avant de connaître un « rattrapage ». Comme nous l'avons indiqué dans la partie précédente, une telle situation est de nature à renchérir le taux de rentabilité exigé pour un investissement (prime de risque), conduisant à des investissements moins importants. La traduction de cette situation en terme de fonction de profit est que des fonctions plus « plates » sont préférables à d'autres plus « pentues ». Par ailleurs, préférer de telles fonctions permet également de limiter la possibilité d'exercer un pouvoir de marché, puisque les profits générés par un retrait volontaire de capacités existantes (en les déclarant indisponibles par exemple) sont alors moindres. C'est cet argument, particulièrement intéressant étant donné l'instabilité des prix à la pointe sur les marchés électriques, qui permet de penser qu'une politique de gestion des réserves opérationnelles couplée à un prix plafond bas est préférable à celle fixant le prix plafond à la valeur du coût de défaillance. Une fonction de profit plus « plate » traduit alors une plus grande stabilité des incitations à investir et crée les conditions pour des primes de risques plus faibles. Cependant, comme le fait remarquer Stoft, une fonction trop « plate » diminue les forces qui tendent à ce que le marché retourne à son équilibre. Une solution serait d'obtenir une fonction assez pentue autour de l'équilibre, mais tendant à s'aplatir quand elle s'en éloigne trop. Il est donc nécessaire de se donner d'autres critères afin de pouvoir sélectionner la combinaison des paramètres politiques la plus à même d'inciter à l'investissement.

Lier gestion des pics de prix et politiques de capacités

L'approche en terme de fonction de profit permet également d'étudier l'ensemble des dispositifs institutionnels qu'on peut regrouper sous l'intitulé « politiques de capacités ». En effet, le constat progressivement réalisé qu'il existe un problème de « revenus manquants » sur les marchés électriques (même si sa reconnaissance sous cet intitulé est très récente) a conduit à un renouvellement d'intérêt pour les dispositifs de rémunération de la capacité indépendants des revenus dégagés par la vente d'énergie (les marchés spot traditionnels). Depuis la rémunération de la capacité intégrée au prix de l'énergie sur le Pool anglais jusqu'aux marchés de capacités à proprement parler, des solutions très différentes ont été mises en place. Elles peuvent bien évidemment se substituer à une politique de gestion des pics de prix. L'approche en terme de fonction de profit a cet avantage de proposer un cadre analytique simple d'évaluation de ces différents dispositifs. Stoft avait déjà mené quelques unes de ces réflexions quand il a proposé cette approche dans son livre « Power system economics » (2002), mais il les a essentiellement approfondies dans le cadre d'un travail sur le design des marchés de capacité pour le NY-ISO (Cramton, Stoft 2006) et par ses activités de consultant. C'est par ce biais que l'approche en terme de fonction de profit a trouvé une reconnaissance institutionnelle, que ce soit dans le design du marché de capacité de New York, ou dans la réflexion de la CPUC sur les marchés de capacité qui utilise explicitement de telles fonctions (Capacity markets white paper 2005).

Mis en place pour la première fois sur le PJM (Pennsylvania Jersey Market, couvrant une zone un peu plus importante que son nom ne l'indique), les marchés de capacité reposent sur un mécanisme d'obligations pour les producteurs à fournir un certain niveau de capacités sous peine de pénalités en cas de non respect de ces obligations (qui sont assorties par des obligations à fournir de l'énergie quand cela est nécessaire). Un objectif global de capacité à respecter est donc retenu par le régulateur et traduit sous forme d'obligations individuelles pour les producteurs.

Un marché de capacité est mis en place afin de pouvoir échanger et valoriser cette nouvelle ressource (nouvelle au sens institutionnel, bien évidemment).

On peut analyser ce dispositif en terme de fonction de profit en introduisant un aléa sur les capacités installées. En effet, un marché de capacité dénué de tout pouvoir de marché et sans possibilité d'échange avec un autre marché voisin, doit alternativement avoir un prix de zéro, si les capacités installées sont supérieures à l'objectif fixé par le régulateur ; et un prix égal à la pénalité si les capacités installées sont inférieures à l'objectif. Considérer les capacités installées comme une variable aléatoire permet d'obtenir une fonction de profit en espérance sans dénaturer le raisonnement qui a mené à sa construction (cela permet également d'obtenir une fonction plus proche du raisonnement d'un investisseur). L'espérance de capacité prend alors le rôle de la variable K dans la partie précédente. On obtient en effet deux situations possibles, chacune affecté d'une probabilité. L'égalité de l'espérance de profit et des coûts fixes définit le niveau d'investissement induit par les revenus dégagés sur le marché de capacité. Une telle approche apparaît tout à fait complémentaire des propositions pour une gestion des pics de prix. Elle permet d'obtenir une fonction de profit relativement pentue autour du point d'équilibre et plus aplatie au fur et à mesure que la capacité installée s'en éloigne.

A bien des égards, l'approche de Stoft permet de renouveler la théorie des marchés d'électricité. Cependant, certaines questions demeurent en suspens : plusieurs arguments laissent à penser que la combinaison d'un marché de capacité avec une gestion de pics de prix réduit la possibilité d'exercice d'un pouvoir de marché, mais aucune preuve formelle n'en est exhibée. Et pour cause, le raisonnement de Stoft autour de la fonction de profit étant mené sans intégrer une telle possibilité. L'égale de la fonction de profit et des coûts fixes suppose qu'il y ait une telle concurrence entre les investisseurs potentiels, que ceux-ci construisent des moyens de production dès qu'il est possible de réaliser un surprofit, même infinitésimal. Le raisonnement de Stoft est assimilable à une relation d'arbitrage sur un marché

financier. Or, les marchés électriques sont des marchés essentiellement oligopolistiques de par la taille des investissements, l'importance des compétences nécessaire à l'opération des centrales ou encore les délais de construction de ces dernières. La partie suivante va essayer de proposer un cadre analytique approprié pour prolonger ces réflexions.

Partie 2 : Le pouvoir de marché par l'investissement

Cette partie a pour objectif d'essayer de résumer dans un cadre théorique les traits principaux de l'investissement en capacités de production, et ainsi de fournir un cadre d'analyse permettant d'étudier les effets des différents dispositifs proposés pour remédier au problème des « revenus manquants ».

L'approche en terme de fonctions de profit a l'avantage de représenter de façon simple les flux de revenus d'un investissement, mais elle laisse de côté certaines problématiques importantes sur les marchés d'électricité. Le modèle proposé dans cette partie utilise les apports de cette approche dans une modélisation plus « explicite » d'un marché électrique. Nous présenterons ainsi un modèle de concurrence imparfaite qui combine en statique l'existence d'un pouvoir de marché avec un processus de fixation des prix simplifié similaire à celui de l'approche en terme de fonction de profit. Le pouvoir de marché existant dans ce modèle ne provient pas de la possibilité pour les producteurs de faire monter les prix en produisant moins (ou en déclarant certaines unités de production indisponibles), mais résulte du processus d'investissement même. En effet, les prix ne dépendent que de la différence entre demande et capacités installées : quand les capacités sont excédentaires, les producteurs ne disposent d'aucun pouvoir de marché et le prix est égal au coût variable du moyen de production. Ce pouvoir de marché sera par la suite appelé pouvoir de marché par l'investissement. Dès qu'il existe un déficit de réserves opérationnelles, le prix prend la valeur du prix plafond. Bien que l'on s'intéresse au cas d'un oligopole, le fait que le prix soit fixé de façon exogène empêche tout exercice d'un pouvoir de marché au sens traditionnel dans les situations de sous capacité (puisque'il est impossible de faire monter le prix en restreignant sa production). En revanche, de par leurs choix d'investissement, les producteurs en oligopole disposent d'un pouvoir sur la distribution de probabilité de prix à l'équilibre. Comme nous allons le voir, cette

distinction a de nombreuses conséquences sur les politiques envisageables pour résoudre le problème des « revenus manquants ».

L'investissement est restreint aux moyens de pointe. Il s'agit d'une simplification, mais proche là aussi de l'esprit de la théorie classique de l'investissement d'un monopole public, car on détermine d'abord le montant de capacité totale à partir du coût de défaillance et des données de coûts d'une unité de pointe. C'est bien la pointe qui permet le bouclage du programme d'investissement (que ce soit dans le cadre d'un monopole public ou dans le cadre d'un secteur déréglementé). Se cantonner à la pointe simplifie les raisonnements tout en conservant le plus possible la portée des résultats.

Ce modèle est assez proche de celui de Crampes et Creti (2005), puisque deux firmes choisissent leurs capacités de production avant de connaître la réalisation de la demande (complètement inélastique), puis connaissent la demande et produisent en tenant compte des contraintes de capacités. Cependant, dans leur modèle le prix de l'électricité produite est déterminé par une enchère, là où le modèle proposé simplifie à l'extrême le processus de fixation des prix afin de pouvoir comparer le niveau d'investissement d'équilibre avec celui d'un monopole public. Meunier et Finon (2006) ont également développé un modèle similaire dans l'esprit, mais s'intéressent au jeu dans lequel interviennent un monopole à la pointe et un gestionnaire de réseau qui peut investir et produire comme menace en cas de sous-investissement. Par ailleurs, ils utilisent une demande élastique. En fait, si ce modèle est à rapprocher d'un autre, c'est surtout celui proposé par Joskow et Tirole dans la partie 4 de leur article de 2006. En effet, dans cette partie la demande est aléatoire et inélastique ; il n'existe qu'un seul type d'unité de production, pour lequel le coût d'investissement est proportionnel aux capacités construites ; enfin, les résultats sur la stratégie d'investissement optimale de second best sont similaires à ceux présentés ici. Cependant, leur modèle ne permet pas de décrire le comportement d'un oligopole, il n'est utilisé que pour obtenir l'optimum de

Ramsey et les moyens de le mettre en oeuvre. Contentons nous de remarquer pour l'instant que l'implémentation qu'ils proposent suppose un prix égal au coût variable lorsque les capacités sont excédentaires et un prix égal au coût de la défaillance lorsque celles-ci sont telles qu'il est nécessaire de recourir à des délestages. Nous aurons l'occasion de revenir sur leurs conclusions à la lumière des résultats du modèle proposé ici.

a-Choix de modélisation et optimum social

Afin de se restreindre aux problèmes de pointe, on modélise la demande résiduelle d'énergie de pointe, c'est à dire la demande non fournie alors que toutes les installations disponibles inférieures dans le « merit order » fonctionnent. On se place donc sur un intervalle de temps réduit à un point : on confronte la demande d'énergie de pointe sur une période (par exemple un an) à l'offre correspondante sur cette période. Ce faisant, on laisse en dehors de la modélisation les problèmes de déficit en puissance éventuels qui peuvent être importants. Mais ceux-ci nécessitent une modélisation beaucoup plus lourde (l'ensemble de la courbe de charge). Cela suppose donc un aléa réduit à la moyenne de la courbe de charge, sans variation de sa forme. On introduit les problèmes de capacités en supposant que chaque producteur doit investir préalablement pour pouvoir produire. Le jeu se fait en deux temps : les firmes investissent d'abord sans connaître la valeur de l'aléa, puis dans la deuxième étape elles connaissent la demande qui leur est adressée et produisent effectivement. La capacité dont il est question dans le modèle n'est alors véritablement qu'une capacité d'énergie, malgré le jeu d'écriture qui consiste à écrire cette capacité en puissance multipliée par la durée de la période. On suit donc ici l'esprit des modèles de Stoft (2002), qui raisonne sur une espérance de demande (par le biais de la durée des pics de prix), à la différence près que l'aléa est explicité dans la fonction de demande.

On suppose que la demande en énergie pour une période t a la forme suivante :

$$D_t = M_t + \varepsilon_t$$

où M_t est non aléatoire et positif ; ε_t suit une loi normale centrée de variance σ^2 .

La demande est donc aléatoire autour d'un terme de « moyenne » ou de tendance M_t .

Cette fonction de demande est une forme particulière de celle de Wolfram (1999) ou de celle de Bushnell et Borenstein (1999) qui modélisent également une concurrence à la Cournot. Mais elle présente deux différences importantes : l'introduction d'un aléa et l'absence de prix dans la fonction. On ne considère pas pour autant une consommation d'électricité inélastique par rapport au prix, même s'il y a lieu de penser que l'élasticité est faible. Mais cette dernière est considérée comme inconnue et se traduit par une fixation exogène du prix. Une telle situation est une radicalisation de la réalité, car si la demande est en effet peu élastique pour les petits consommateurs (sauf tarifs particuliers comme les EJP en France), les plus gros consommateurs (industriels) qui peuvent arrêter temporairement leur production sont susceptibles de réduire leur consommation contre rémunération. Cependant, il est plus facile en général de considérer ces marges de manœuvre comme une réserve supplémentaire du côté de l'offre gérée sur valeur d'usage, d'où la forme retenue de la demande. Cette fonction de demande représente ce que Stoft (2002) appelle « augmented load », c'est-à-dire correspond à la demande à laquelle est ajoutée la puissance nécessaire pour compenser les centrales connaissant une avarie. On peut considérer que cette fonction de demande inclue les besoins en réserves opérationnelles dans le terme M_i .

La forme de l'aléa affectant la demande en énergie est extrêmement simple par rapport à la réalité. En particulier, il existe une faible auto-corrélation de la demande de puissance dans le temps. Et, plus important, il y a tout lieu de penser que la demande d'énergie résiduelle n'est pas affectée par des chocs symétriques, mais plutôt par des chocs de moyenne négative. Pour autant, et en l'absence d'une caractérisation générale de la fonction de demande (qui varie de pays en pays), il est plus simple de faire la simplification d'une demande suivant une loi normale. Car la plupart des résultats présentés par la suite le sont dans un cadre statique (ce qui élimine les problèmes d'auto corrélation), et considérer des chocs symétriques ne facilite pas la démonstration, puisque cela augmente l'espérance de gain d'un investisseur. Par ailleurs, les résultats peuvent être étendus à d'autres lois de probabilité sous certaines conditions.

On suppose que deux entreprises se font concurrence pour la fourniture de la demande d'électricité. Le problème est celui d'une décision en avenir incertain, l'investisseur potentiel fonde donc sa décision sur son espérance de gain. On suppose des agents neutres au risque, par commodité analytique, mais cela a plutôt tendance à sous estimer les problèmes de sous investissement. La modélisation du duopole adoptée est un mélange d'équilibre de Bertrand, lorsqu'il n'y a pas de contraintes de capacités, et d'équilibre de Cournot simplifié, lorsque celles-ci existent. Comme nous l'avons vu, les SFE sont encadrées par les équilibres de Bertrand et Cournot. Ces derniers représentent les cas limites des équilibres des modèles de la littérature. Une telle approche, de plus, s'inspire de l'esprit de l'article de Kreps et Scheinkman (1983) où une concurrence en prix avec des capacités limitées (qui doivent également être construites préalablement à la production) permet d'obtenir un équilibre de Cournot. Mais leur modèle n'est pas repris puisqu'il ne permet pas d'obtenir des équilibres en stratégies pures lorsqu'il y a des surcapacités.

Le processus de fixation des prix a été simplifié en posant que, lorsque la demande est inférieure aux capacités installées, les entreprises sont soumises à une concurrence à la Bertrand et le prix est égal au coût variable de production des entreprises. Dans le cas inverse, des achats sont nécessaires pour restaurer les réserves opérationnelles, et le prix est donc égal au prix plafond du marché. Ce cadre est donc similaire à celui de l'approche en terme de fonction de profit, où la rente de rareté est obtenue lorsque la confrontation demande / capacités déclenche un pic de prix. Cependant, le fait d'explicitement la décision d'investissement permet de combiner dans un même modèle l'existence d'un pouvoir de marché (puisque les entreprises du duopole choisissent le niveau des capacités en fonction des revenus qu'elles vont dégager) et l'analyse des pics de prix. Notons enfin que le modèle se généralise sans problème au cas d'un n-oligopole (comme il est démontré plus loin), mais que pour la clarté de la démonstration le duopole a été retenu.

Enfin, reste à savoir comment confronter le niveau d'investissement total obtenu par le duopole au niveau optimal de capacité, tel que défini dans la première partie. Etant donné les choix de modélisation précédemment exposés, le niveau de capacité optimal va être donné par la résolution du programme d'investissement d'un monopole public confronté aux mêmes conditions de demande et de coûts que les entreprises du duopole. Ce programme va avoir pour critère la minimisation des coûts de fourniture de la demande, sachant que toute demande non satisfaite est valorisée au coût de défaillance. Les résultats correspondent à l'optimum de Ramsey. On exploite ici l'analogie qui existe entre la théorie de Boiteux (Varoquaux 1996) quant à l'investissement d'un monopole électrique public et l'approche en terme de fonction de profit.

De la comparaison entre les niveaux d'investissement du duopole et du monopole, on peut conclure ou non à l'existence d'un phénomène de sous investissement du duopole pour différents niveaux du prix plafond. En particulier, il est intéressant de comparer le niveau d'investissement du monopole public avec celui du duopole dans le cas où le prix plafond est égal au coût de défaillance (cas de « VOLL pricing »). De même, la question se pose de savoir si le pouvoir de marché exercé par les entreprises invalide la capacité d'une politique de gestion des pics de prix à restaurer les incitations à investir. Ce modèle peut donc permettre d'approfondir l'analyse de Stoft et de faire le lien avec les modélisations plus « traditionnelles » des marchés électriques.

i. Le programme d'investissement d'un monopole public

Dans le cas d'un monopole public l'énergie vendue est tarifée de façon exogène. Le problème est alors de minimiser les coûts de fourniture de la demande, sachant que tout mégawatheure non fourni coûte le coût de la défaillance C_D , qu'un mégawatt de moyen de pointe coûte I et que la production d'un mégawatheure a un coût variable C_P . On considère que le processus de décision se fait en deux étapes : le monopole décide d'abord d'investir pour construire une quantité \bar{Q} sans connaître la réalisation de la variable aléatoire ε , il ne fonde donc sa décision que sur son espérance de coût ; puis le niveau de la demande est connu et le monopole produit l'offre correspondante jusqu'à saturation de ses capacités auquel cas la demande non fournie est considérée comme de la défaillance. Pour simplifier les notations, on écrit \bar{Q} la capacité de production en mégawatt multipliée par la durée de la période sur laquelle on s'intéresse à l'équilibre offre-demande. On peut ainsi traiter puissances et énergies de façon quasi-équivalente, puisque \bar{Q} représente alors la capacité en énergie sur la période. L'investissement du monopole s'effectue donc au niveau de capacité qui minimise l'espérance de coût de fourniture de la demande (ce qui revient à maximiser le surplus social), ce qui peut s'écrire :

$$\min_{\bar{Q}} E[\text{coût}] = \min_{\bar{Q}} E \left[I\bar{Q} + C_P D 1_{\{D < \bar{Q}\}} + (C_P \bar{Q} + C_D (D - \bar{Q})) 1_{\{D \geq \bar{Q}\}} \right]$$

où $1_{\{D < \bar{Q}\}}$ est la fonction indicatrice de l'ensemble $\{D < \bar{Q}\}$

Ce problème peut se simplifier en notant qu'il est équivalent à :

$$\min_{\bar{Q}} I\bar{Q} + E \left[C_P D 1_{\{D < \bar{Q}\}} + (C_P \bar{Q} + C_D (D - \bar{Q})) 1_{\{D \geq \bar{Q}\}} \right]$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_P E \left[D 1_{\{D < \bar{Q}\}} + \left(\bar{Q} + \frac{C_D}{C_P} (D - \bar{Q}) \right) 1_{\{D \geq \bar{Q}\}} \right]$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_P E \left[(M + \varepsilon) 1_{\{\varepsilon < \bar{Q} - M\}} + \left(\left(1 - \frac{C_D}{C_P} \right) \bar{Q} + \frac{C_D}{C_P} D \right) 1_{\{\varepsilon < \bar{Q} - M\}} \right]$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_P E \left[(M + \varepsilon) 1_{\{\varepsilon < \bar{Q} - M\}} + \left(\left(1 - \frac{C_D}{C_P} \right) \bar{Q} + \frac{C_D}{C_P} D \right) (1 - 1_{\{\varepsilon < \bar{Q} - M\}}) \right]$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_P E \left[\begin{aligned} & (M + \varepsilon) 1_{\{\varepsilon < \bar{Q} - M\}} + \left(\left(1 - \frac{C_D}{C_P} \right) \bar{Q} + \frac{C_D}{C_P} (M + \varepsilon) \right) \\ & - \left(\left(1 - \frac{C_D}{C_P} \right) \bar{Q} + \frac{C_D}{C_P} (M + \varepsilon) \right) 1_{\{\varepsilon < \bar{Q} - M\}} \end{aligned} \right]$$

Par linéarité de l'espérance, on obtient :

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_D (M + E(\varepsilon)) + (C_P - C_D) \bar{Q} \\ + C_P E \left[(M + \varepsilon) 1_{\{\varepsilon < \bar{Q} - M\}} - \left(\left(1 - \frac{C_D}{C_P} \right) \bar{Q} + \frac{C_D}{C_P} (M + \varepsilon) \right) 1_{\{\varepsilon < \bar{Q} - M\}} \right]$$

En se souvenant que ε suit une loi normale centrée de variance σ^2 et donc que l'espérance de ε vaut 0, on peut simplifier l'équivalence :

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_D M + (C_P - C_D) \bar{Q} \\ + C_P E \left[\left(1 - \frac{C_D}{C_P} \right) (M + \varepsilon) 1_{\{\varepsilon < \bar{Q} - M\}} - \left(1 - \frac{C_D}{C_P} \right) \bar{Q} 1_{\{\varepsilon < \bar{Q} - M\}} \right]$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_D M + (C_P - C_D) \bar{Q} + (C_P - C_D) M E\left(1_{\{\varepsilon < \bar{Q} - M\}}\right) + (C_P - C_D) E\left(\varepsilon 1_{\{\varepsilon < \bar{Q} - M\}}\right) - (C_P - C_D) \bar{Q} E\left(1_{\{\varepsilon < \bar{Q} - M\}}\right)$$

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_D M + (C_P - C_D) \bar{Q} + (C_P - C_D) M \int_{-\infty}^{\bar{Q}-M} f(t) dt + (C_P - C_D) \int_{-\infty}^{\bar{Q}-M} t f(t) dt - (C_P - C_D) \bar{Q} \int_{-\infty}^{\bar{Q}-M} f(t) dt$$

où $f(t)$ est la densité de la variable aléatoire ε qui suit une loi normale.

On a donc :

$$\Leftrightarrow \min_{\bar{Q}} I\bar{Q} + C_D M + (C_P - C_D) \bar{Q} + (C_P - C_D) M \int_{-\infty}^{\bar{Q}-M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt + (C_P - C_D) \int_{-\infty}^{\bar{Q}-M} t \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt - (C_P - C_D) \bar{Q} \int_{-\infty}^{\bar{Q}-M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt$$

Les intégrales ne sont évidemment pas explicitables, de même que les probabilités d'une variable normale. Par contre, les fonctions de la variable \bar{Q} qu'elles constituent sont dérivables. On peut donc bien obtenir les conditions au premier et au second ordre qu'une valeur de \bar{Q} minimisant l'expression ci-dessus doit respecter.

Condition du premier ordre :

$$\frac{\partial E(\text{coût})}{\partial \bar{Q}} = I + (C_P - C_D) + (C_P - C_D) M \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}} + (C_P - C_D) (\bar{Q} - M) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}} - (C_P - C_D) \bar{Q} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}} - (C_P - C_D) \int_{-\infty}^{\bar{Q}-M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt = 0$$

Ce qui se simplifie en :

$$\frac{\partial E(\text{coût})}{\partial \bar{Q}} = I + (C_P - C_D) - (C_P - C_D) \int_{-\infty}^{\bar{Q}-M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt = 0$$

L'intégrale, non plus, n'est pas ici explicitable, mais permet une réécriture de l'égalité :

$$1 - \frac{I}{(C_D - C_P)} = P(\varepsilon < \bar{Q} - M) = P(D < \bar{Q})$$

On obtient donc une relation entre les différents coûts et la probabilité qu'il n'y ait pas défaillance, relation qui permet d'obtenir implicitement la capacité optimale \bar{Q}_M^* à construire par le monopole public. Pour que cette équation ait un sens, il faut aussi que la condition suivante soit respectée :

$$\frac{I}{(C_D - C_P)} < 1$$

On supposera que c'est le cas, mais il est possible de relâcher cette condition. On peut noter que cette dernière permet de faire le lien entre coût d'investissement et coût de la défaillance, et donc éventuellement de définir celui-ci à partir d'un critère physique.

L'égalité précédente permet également de connaître l'incidence (en terme de sens de variation) des variables exogènes sur \bar{Q}_M^* .

Il faut également vérifier que la condition de deuxième ordre est bien respectée. Celle-ci est égale à :

$$\frac{\partial^2 E(\text{coût})}{\partial \bar{Q}^2} = -(C_P - C_D) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}}$$

La dérivée seconde de l'espérance de coût est négative, le niveau d'investissement d'équilibre est donc bien un minimum.

ii. Analyse du niveau optimal de capacité

Ainsi, puisque $P(D < \bar{Q})$ est une fonction croissante de \bar{Q} :

- une hausse de C_D implique une variation positive de \bar{Q}_M^* ;
- une hausse de C_P implique une variation négative de \bar{Q}_M^* , le différentiel entre le coût de la défaillance et le coût proportionnel diminue, rendant le recours à la défaillance moins coûteux relativement à la production en pointe ;
- une hausse de I implique une variation négative de \bar{Q}_M^* , là aussi c'est le coût de la production en pointe qui augmente relativement à la non production ;
- une hausse de M implique une variation positive de \bar{Q}_M^* , afin de maintenir constante la probabilité qu'il n'y ait pas défaillance ;
- une hausse de σ^2 implique une variation positive de \bar{Q}_M^* lorsque $\bar{Q}_M^* - M$ est négatif, et implique une variation négative dans le cas inverse. Dans le premier cas, une hausse de la variance se traduit par une hausse de l'espérance de coût de défaillance, puisque l'espérance physique de défaillance est positive. Dans le second cas, l'espérance physique de défaillance est négative, donc une hausse de la variance se traduit par une baisse de l'espérance du coût de défaillance.

Si on appelle A , la probabilité $P(D < \bar{Q})$, ce dernier point est démontrable en remarquant que :

$$\frac{\partial A}{\partial \sigma^2} = \frac{\partial \int_{-\infty}^{\bar{Q}-M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt}{\partial \sigma^2}$$

$$\frac{\partial A}{\partial \sigma^2} = -\frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\bar{Q}-M} \frac{-4\sigma t^2}{4\sigma^4} e^{-\frac{t^2}{2\sigma^2}} dt$$

$$\frac{\partial A}{\partial \sigma^2} = -\frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt + \frac{1}{\sqrt{2\pi}\sigma^4} \int_{-\infty}^{\bar{Q}-M} t^2 e^{-\frac{t^2}{2\sigma^2}} dt$$

$$\frac{\partial A}{\partial \sigma^2} = -\frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt + \frac{1}{\sqrt{2\pi}\sigma^4} \left[\left[-\sigma^2 t e^{-\frac{t^2}{2\sigma^2}} \right]_{-\infty}^{\bar{Q}-M} + \sigma^2 \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt \right]$$

$$\frac{\partial A}{\partial \sigma^2} = -\frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt + \frac{1}{\sqrt{2\pi}\sigma^4} \left[-\sigma^2 (\bar{Q}-M) e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}} \right] + \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\bar{Q}-M} e^{-\frac{t^2}{2\sigma^2}} dt$$

$$\frac{\partial A}{\partial \sigma^2} = \frac{-1}{\sqrt{2\pi}\sigma^2} (\bar{Q}-M) e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}}$$

$$\frac{\partial A}{\partial \sigma^2} < 0, \text{ si } \bar{Q} - M < 0 \text{ et inversement.}$$

Or, de par la définition de \bar{Q}_M^* , $\bar{Q}_M^* - M$ est positif. Donc une hausse de σ^2 implique bien une variation positive de \bar{Q}_M^* .

b-Le duopole déréglementé

Le cas d'un marché déréglementé avec des firmes privées en concurrence est représenté par un duopole, où l'on reprend les notations et les variables du monopole. On considère donc que deux firmes (indiquées i et j) sont en concurrence sur le marché de pointe.

La séquence des décisions du jeu est la même que pour le monopole : les deux firmes décident dans un premier temps de leurs capacités de production, sans connaître la réalisation de la demande, puis dans un deuxième temps la demande est connue et les firmes produisent et vendent leur production. Les coûts sont identiques à ceux du monopole, à la différence près qu'une demande non fournie n'a pas de coût pour une firme.

Dans le deuxième temps du jeu on suppose que, lorsque la demande est inférieure à la somme des capacités de production, la concurrence entre firmes s'opère par les prix. Le prix est alors égal au coût marginal de production, soit au coût proportionnel C_p (équilibre de Bertrand). On suppose, en revanche, que, lorsque la demande est supérieure à la somme des capacités, la demande excédentaire est réduite à zéro par une élévation du niveau des prix à \bar{P} , qui correspond à la valeur du prix plafond sur le marché. Cette situation correspond à un pic de prix. On va résoudre le modèle en supposant que le niveau de réserves opérationnelles exigé est égal à zéro. On peut donc éviter d'introduire une variable supplémentaire correspondant à ce niveau.

Notons également que la demande non fournie n'est pas intégrée comme coût pour les firmes, contrairement au cas du monopole public, qui paye la défaillance.

On obtient deux cas possibles suivant les réalisations de ε : une situation de sur-capacité avec concurrence en prix où le prix d'équilibre est égal au coût proportionnel, chaque firme est donc indifférente à son niveau de production ; et

une situation de sous-capacité qui se règle par élévation du prix d'équilibre à un niveau exogène et où chaque firme produit à pleine capacité.

Les firmes vont donc choisir leurs niveaux de capacité sur la base de leurs espérances de gain (puisque'elles sont neutres au risque). On cherche alors l'équilibre de Cournot-Nash défini par l'intersection de leurs fonctions de réaction.

i. Résolution du programme d'investissement d'une firme

L'espérance de profit de la firme i s'écrit alors :

$$\max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} E \left[(C_P \bar{Q}_i - C_P \bar{Q}_i - I \bar{Q}_i) 1_{\{D < \bar{Q}_i + \bar{Q}_j\}} + (\bar{P} \bar{Q}_i - C_P \bar{Q}_i - I \bar{Q}_i) 1_{\{D \geq \bar{Q}_i + \bar{Q}_j\}} \right]$$

Ce problème peut se simplifier en notant qu'il est équivalent à :

$$\max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} \left\{ -I \bar{Q}_i + E \left[(\bar{P} \bar{Q}_i - C_P \bar{Q}_i) 1_{\{D \geq \bar{Q}_i + \bar{Q}_j\}} \right] \right\}$$

$$\Leftrightarrow \max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} \left\{ -I \bar{Q}_i + E \left[(\bar{P} \bar{Q}_i - C_P \bar{Q}_i) (1 - 1_{\{D < \bar{Q}_i + \bar{Q}_j\}}) \right] \right\}$$

$$\Leftrightarrow \max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} \left\{ -I \bar{Q}_i + \bar{Q}_i (\bar{P} - C_P) - (\bar{P} \bar{Q}_i - C_P \bar{Q}_i) E[1_{\{D < \bar{Q}_i + \bar{Q}_j\}}] \right\}$$

$$\Leftrightarrow \max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} \left\{ -I \bar{Q}_i + \bar{Q}_i (\bar{P} - C_P) - \bar{Q}_i (\bar{P} - C_P) P(D < \bar{Q}_i + \bar{Q}_j) \right\}$$

On a également les égalités suivantes, en appelant A la probabilité qui intervient dans l'équation précédente :

$$A = P(D < \bar{Q}_i + \bar{Q}_j) = P(\varepsilon < \bar{Q}_i + \bar{Q}_j - M) = \int_{-\infty}^{\bar{Q}_i + \bar{Q}_j - M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt$$

A est donc une fonction dérivable de \bar{Q}_i . Il est dès lors possible d'écrire la condition de premier ordre de la maximisation :

$$\frac{\partial E[\pi_i]}{\partial \bar{Q}_i} = -I + (\bar{P} - C_p) - \bar{Q}_i(\bar{P} - C_p) \frac{\partial A}{\partial \bar{Q}_i} - (\bar{P} - C_p)A = 0$$

$$\Leftrightarrow -\frac{I}{(\bar{P} - C_p)} + 1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A = 0$$

$$\boxed{\Leftrightarrow 1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A = \frac{I}{(\bar{P} - C_p)}}$$

La fonction de réaction de la firme i est définie implicitement par l'égalité précédente. En faisant la différence avec celle de la firme j on obtient l'égalité :

$$-\bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A + \bar{Q}_j \frac{\partial A}{\partial \bar{Q}_j} + A = 0$$

$$\Leftrightarrow -\bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} + \bar{Q}_j \frac{\partial A}{\partial \bar{Q}_j} = 0$$

Or, la dérivée partielle de A par rapport à \bar{Q}_i est égale à la dérivée partielle de A par rapport à \bar{Q}_j , comme on peut le montrer :

$$\begin{aligned} \frac{\partial A}{\partial \bar{Q}_i} &= \frac{\partial P(\varepsilon < \bar{Q}_i + \bar{Q}_j - M)}{\partial \bar{Q}_i} = \frac{\partial \int_{-\infty}^{\bar{Q}_i + \bar{Q}_j - M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt}{\partial \bar{Q}_i} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \\ &= \frac{\partial \int_{-\infty}^{\bar{Q}_i + \bar{Q}_j - M} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} dt}{\partial \bar{Q}_j} = \frac{\partial P(\varepsilon < \bar{Q}_i + \bar{Q}_j - M)}{\partial \bar{Q}_j} = \frac{\partial A}{\partial \bar{Q}_j} \end{aligned}$$

On a donc :

$$\frac{\partial A}{\partial \bar{Q}_i} (\bar{Q}_j - \bar{Q}_i) = 0$$

Comme $\frac{\partial A}{\partial \bar{Q}_i} > 0$, du fait de l'exponentielle, on a finalement :

$$\bar{Q}_j = \bar{Q}_i$$

L'équilibre de Cournot-Nash est donc symétrique. Il est alors donné par l'égalité suivante :

$$1 - \bar{Q} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2\bar{Q}-M)^2}{2\sigma^2}} - P(D < 2\bar{Q}) = \frac{I}{(\bar{P} - C_P)} \quad (1)$$

Ce qui peut également s'écrire sous la forme :

$$1 - \bar{Q} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2\bar{Q}-M)^2}{2\sigma^2}} - P(D < 2\bar{Q}) - \frac{I}{(\bar{P} - C_P)} = 0$$

Soit :

$$1 - \bar{Q} \frac{\partial P(D < 2\bar{Q})}{\partial \bar{Q}} - P(D < 2\bar{Q}) = \frac{I}{(\bar{P} - C_P)}$$

On note \bar{Q}^* , la valeur à l'équilibre, qui est positive. La formule obtenue est semblable à celle du monopole, mais avec \bar{P} à la place de C_D et la dérivée de la probabilité qui vient s'intercaler.

Unicité et condition du second ordre

Il faut vérifier qu'il s'agit bien d'un maximum. La condition du deuxième ordre s'écrit :

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = -\bar{Q}_i \frac{\partial^2 A}{\partial \bar{Q}_i^2} - 2 \frac{\partial A}{\partial \bar{Q}_i}$$

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = \frac{1}{\sigma^2} \bar{Q}_i (\bar{Q}_i + \bar{Q}_j - M) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} - 2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}}$$

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \left(\frac{1}{\sigma^2} \bar{Q}_i (\bar{Q}_i + \bar{Q}_j - M) - 2 \right)$$

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} (\bar{Q}_i^2 + (\bar{Q}_j - M)\bar{Q}_i - 2\sigma^2)$$

Le signe de la dérivée seconde du profit de i est alors donné par le signe du polynôme $\bar{Q}_i^2 + (\bar{Q}_j - M)\bar{Q}_i - 2\sigma^2$. Si on se place à l'équilibre symétrique, ce polynôme se réécrit :

$$\bar{Q}^{*2} + (\bar{Q}^* - M)\bar{Q}^* - 2\sigma^2 = 2\bar{Q}^{*2} - M\bar{Q}^* - 2\sigma^2$$

Son discriminant Δ s'écrit :

$$\Delta = M^2 + 16\sigma^2 > 0$$

Les racines du polynôme sont réelles et ont pour valeurs :

$$r_1 = \frac{M + \sqrt{M^2 + 16\sigma^2}}{4} > 0 \quad r_2 = \frac{M - \sqrt{M^2 + 16\sigma^2}}{4} < 0$$

Pour une valeur nulle de \bar{Q}^* , le polynôme est négatif ; donc sur l'intervalle $[r_2, r_1]$, la dérivée seconde est négative et \bar{Q}^* est un maximum local. Pour des valeurs de \bar{Q}^* supérieures à r_1 , la dérivée seconde est positive et \bar{Q}^* est un minimum local. La valeur de \bar{Q}^* déterminée par la condition de premier ordre est-elle unique ? Pour répondre à cette question, il faut savoir si la fonction de réaction de la firme i est monotone sur $[r_2, r_1]$. Or, nous venons de calculer sa dérivée par rapport à \bar{Q}_i , qui est négative sur $[r_2, r_1]$. Elle est donc décroissante et monotone sur cet intervalle et réalise une bijection de $[r_2, r_1]$ sur l'intervalle suivant :

$$\left[1 - r_1 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2r_1-M)^2}{2\sigma^2}} - P(D < 2r_1) - \frac{I}{(\bar{P} - C_p)}, 1 - r_2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2r_2-M)^2}{2\sigma^2}} - P(D < 2r_2) - \frac{I}{(\bar{P} - C_p)} \right]$$

Nous pouvons en conclure que la solution de l'équation (1) est unique et existe à condition que les inégalités suivantes soient vérifiées simultanément :

$$1 - r_1 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2r_1-M)^2}{2\sigma^2}} - P(D < 2r_1) - \frac{I}{(\bar{P} - C_p)} < 0$$

$$1 - r_2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2r_2-M)^2}{2\sigma^2}} - P(D < 2r_2) - \frac{I}{(\bar{P} - C_p)} > 0$$

Enfin, \bar{Q}^* est positive si la condition suivante est vérifiée :

$$1 - 0 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(0-M)^2}{2\sigma^2}} - P(D < 0) - \frac{I}{(\bar{P} - C_p)} > 0$$

En effet, dans ce cas, l'équation (1) qui détermine le niveau d'investissement optimal d'une firme prend ses valeurs sur l'intervalle $[0, r_1]$. Les conditions qui doivent être remplies pour qu'une solution positive unique existe sur $[0, r_1]$ sont donc :

$$1 - 0 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(0-M)^2}{2\sigma^2}} - P(D < 0) - \frac{I}{(\bar{P} - C_p)} > 0$$

$$1 - r_1 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(2r_1-M)^2}{2\sigma^2}} - P(D < 2r_1) - \frac{I}{(\bar{P} - C_p)} < 0$$

La première peut se réécrire sous la forme :

$$1 - P(D < 0) - \frac{I}{(\bar{P} - C_p)} > 0$$

Etant donné qu'une probabilité est toujours inférieure ou égale à 1, une condition nécessaire (mais non suffisante) pour obtenir cette inégalité est que :

$$\frac{I}{(\bar{P} - C_p)} < 1$$

On retrouve une condition d'existence de la solution équivalente à celle du monopole public. Cependant, la deuxième inégalité est un peu plus complexe puisqu'elle peut s'écrire :

$$1 - \frac{M + \sqrt{M^2 + 16\sigma^2}}{4} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\frac{-2M + \sqrt{M^2 + 16\sigma^2}}{4})^2}{2\sigma^2}} - P(D < \frac{M + \sqrt{M^2 + 16\sigma^2}}{2}) - \frac{I}{(\bar{P} - C_p)} < 0$$

Comme les deux inégalités doivent être vérifiées simultanément, le fait que le rapport des coûts doive être inférieur à un, rend la vérification de la seconde plus délicate. Elle semble « évidemment » inférieure à zéro, puisque la probabilité dans l'expression est proche de un et qu'elle vient se soustraire au seul terme positif qui vaut un. Cependant aucune preuve formelle ne peut en être apportée. Nous supposons donc cette égalité validée.

Ces inégalités affectent les rapports possibles entre le coût d'investissement et la rente de rareté dégagée par la vente d'une unité d'énergie pendant un pic de prix.

D'autres inégalités pèsent sur \bar{Q}^* :

$$0 < \bar{Q}^* < r_1$$

Le fait que \bar{Q}^* doive être inférieur à r_1 peut se réécrire sous la forme :

$$\bar{Q}^* < \frac{M + \sqrt{M^2 + 16\sigma^2}}{4}$$

La solution du modèle (pour une firme) n'est définie que sur un intervalle dont la borne supérieure est de l'ordre de la moitié de la moyenne de la demande plus une fois l'écart type. Par conséquent, le total des capacités doit être grossièrement inférieur à la moyenne de la demande plus deux fois l'écart type. Cet intervalle capture une part importante des variations possibles de la demande, il peut donc être estimé raisonnable. Nous reviendrons plus loin sur ces conditions. Contentons-

nous de considérer qu'elles sont remplies par les différents paramètres, de telle sorte qu'une solution unique existe.

ii. Analyse de la solution

Il est maintenant possible d'essayer de caractériser la solution \bar{Q}^* relativement à des variations des variables exogènes et surtout de la comparer à celle du monopole. Si on note \bar{Q}_D^* la solution du duopole et \bar{Q}_M^* celle du monopole, on a les équations suivantes :

$$1 - \bar{Q}_D^* \frac{\partial P(D < 2\bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*) - P(D < 2\bar{Q}_D^*) = \frac{I}{(\bar{P} - C_P)}$$

$$1 - P(D < \bar{Q}_M^*) = \frac{I}{(C_D - C_P)}$$

Comme C_D est supérieur à \bar{P} , on a l'inégalité :

$$\frac{I}{(\bar{P} - C_P)} > \frac{I}{(C_D - C_P)}$$

De plus, $1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A$ est une fonction décroissante de \bar{Q} , donc \bar{Q}_D^* est à son maximum lorsque \bar{P} est égal à C_D . Or, la valeur de \bar{Q}_D^* lorsque cela est le cas est donnée par :

$$1 - \bar{Q}_D^* \frac{\partial P(D < 2\bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*) - P(D < 2\bar{Q}_D^*) = \frac{I}{(C_D - C_P)}$$

Ce qui peut se réécrire sous la forme :

$$1 - P(D < 2 \bar{Q}_D^*) = \frac{I}{(C_D - C_P)} + \bar{Q}_D^* \frac{\partial P(D < 2 \bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*)$$

Soit, aussi :

$$1 - P(D < 2 \bar{Q}_D^*) = 1 - P(D < \bar{Q}_M^*) + \bar{Q}_D^* \frac{\partial P(D < 2 \bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*)$$

Or, on a l'inégalité :

$$\bar{Q}_D^* \frac{\partial P(D < 2 \bar{Q}_D^*)}{\partial \bar{Q}_i}(\bar{Q}_D^*) > 0$$

Et la fonction qui à un niveau de capacité \bar{Q} associe la probabilité $1 - P(D < 2 \bar{Q})$ est décroissante en \bar{Q} .

Donc le niveau maximal de capacité de deux fois \bar{Q}_D^* est inférieur à \bar{Q}_M^* . La capacité installée totale de production sous duopole est toujours inférieure à celle sous monopole public, quel que soit le niveau de prix que permet une sous capacité. C'est un résultat assez fort, mais qui n'est valide que dans la limite des intervalles de valeurs qui permettent ces conclusions, dont nous avons vu qu'ils étaient « raisonnables ».

Ainsi, puisque $1 - \bar{Q} \frac{\partial A}{\partial \bar{Q}_i}(\bar{Q}) - A(\bar{Q})$ est une fonction décroissante de \bar{Q} , on a les relations suivantes entre \bar{Q}_D^* et les variables exogènes du modèle :

- une hausse de \bar{P} implique une variation positive de \bar{Q}_D^* ;

- une hausse de C_p implique une variation négative de \bar{Q}_D^* , le différentiel entre le prix lorsqu'il y a sous capacité et le coût proportionnel diminue, rendant les situations de sous capacité moins rentables ;
- une hausse de I implique une variation négative de \bar{Q}_D^* , le coût de la production en pointe augmente, abaissant la rentabilité d'investir ;
- une hausse de M implique une variation positive de \bar{Q}_D^* , afin de maintenir constante la probabilité qu'il n'y ait pas défaillance ;
- une hausse de σ^2 implique une variation positive de \bar{Q}_D^* dans la plupart des cas.

En effet :

$$\frac{\partial A}{\partial \sigma^2} = \frac{-1}{\sqrt{2\pi}\sigma^2} (\bar{Q} - M) e^{-\frac{(\bar{Q}-M)^2}{2\sigma^2}}$$

$$\begin{aligned} \frac{\partial (1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A)}{\partial \sigma^2} &= -\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \bar{Q}_i + \frac{1}{\sqrt{2\pi}\sigma} \frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{\sigma^3} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \bar{Q}_i \\ &- \frac{1}{\sqrt{2\pi}\sigma^2} (\bar{Q}_i + \bar{Q}_j - M) e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \end{aligned}$$

$$\frac{\partial (1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A)}{\partial \sigma^2} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \left[-\bar{Q}_i + \frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{\sigma^2} \bar{Q}_i - (\bar{Q}_i + \bar{Q}_j - M) \right]$$

$$\frac{\partial (1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A)}{\partial \sigma^2} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{2\sigma^2}} \left[-\bar{Q}_i + \frac{(\bar{Q}_i + \bar{Q}_j - M)^2}{\sigma^2} \bar{Q}_i - (\bar{Q}_i + \bar{Q}_j - M) \right]$$

En se plaçant à l'équilibre symétrique, cette égalité devient :

$$\frac{\partial (1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A)}{\partial \sigma^2} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(2\bar{Q}_D^* - M)^2}{2\sigma^2}} \left[-\bar{Q}_D^* + \frac{(2\bar{Q}_D^* - M)^2}{\sigma^2} \bar{Q}_D^* - (2\bar{Q}_D^* - M) \right]$$

$$\frac{\partial (1 - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} - A)}{\partial \sigma^2} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(2\bar{Q}_D^* - M)^2}{2\sigma^2}} \left[4\bar{Q}_D^{*3} - 4M\bar{Q}_D^{*2} + M^2\bar{Q}_D^* - 3\sigma^2\bar{Q}_D^* + \sigma^2 M \right]$$

Le signe de cette dérivée dépend donc du signe du polynôme en \bar{Q}_D^* , P :

$$P(\bar{Q}_D^*) = 4\bar{Q}_D^{*3} - 4M\bar{Q}_D^{*2} + (M^2 - 3\sigma^2)\bar{Q}_D^* + \sigma^2 M.$$

Sans résoudre explicitement l'équation $P(\bar{Q}_D^*) = 0$ il est possible de connaître son signe en raisonnant sur l'intervalle :

$$\left[0, \frac{M + \sqrt{(\bar{Q}_D^* - M)^2 + 8\sigma^2}}{2} \right]$$

En effet, \bar{Q}_D^* doit être inférieur au deuxième terme du fait de l'inégalité :

$2\bar{Q}_D^* \leq M + \sqrt{(\bar{Q}_D^* - M)^2 + 8\sigma^2}$. Or l'équation $P(\bar{Q}_D^*) = 0$ peut se réécrire sous la forme :

$$\bar{Q}_D^* (4\bar{Q}_D^{*2} - 4M\bar{Q}_D^* + (M^2 - 3\sigma^2)) + \sigma^2 M = 0$$

Le polynôme de degré 2 qui apparaît ainsi a pour racines :

$$r_1 = \frac{M}{2} - \frac{\sqrt{3}}{8}\sigma \quad r_2 = \frac{M}{2} + \frac{\sqrt{3}}{8}\sigma$$

Sur l'intervalle $\left[r_2, \frac{M}{2} + \frac{\sqrt{M^2 + 8\sigma^2}}{2} \right]$ P est donc positif puisque le polynôme de degré 2 l'est et que \bar{Q}_D^* est toujours positif. Sur l'intervalle $[r_1, r_2]$, qui inclut 0, le signe de P dépend de la valeur de \bar{Q}_D^* . En fait, sur l'intervalle compris entre la racine de P comprise entre r_1 et r_2 , et r_2 , P reste positif. En dessous de la racine de P évoquée, P devient négatif.

Le sens de variation de \bar{Q}_D^* dans le cas d'une hausse de σ^2 est donc positif sur un large intervalle de valeurs de \bar{Q}_D^* , mais il existe un intervalle pour lequel ce sens est inversé.

iii. Extension et généralisation du modèle

Le modèle exposé précédemment peut être généralisé au cas d'un n-oligopole et à d'autres distributions de probabilité pour la demande que la loi normale. De même, s'il n'admet de solution que sur un intervalle de valeurs restreint, ce dernier peut être étendu en considérant plusieurs périodes de jeu. Ces différentes extensions (et les conditions attenantes) sont examinées ici.

Relâchement des contraintes

Une des limites du modèle réside dans la restriction qu'il impose aux paramètres de coûts afin d'avoir une solution dans le cas du duopole. Nous avons vu que le rapport des coûts devait respecter les inégalités suivantes :

$$1 - \frac{M + \sqrt{M^2 + 16\sigma^2}}{4} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(-2M + \sqrt{M^2 + 16\sigma^2})^2}{4\sigma^2}} - P(D < \frac{M + \sqrt{M^2 + 16\sigma^2}}{2}) < \frac{I}{(\bar{P} - C_p)}$$
$$\frac{I}{(\bar{P} - C_p)} < 1 - P(D < 0)$$

Le terme de gauche de la première inégalité est soit négatif (auquel cas il n'y a pas de contrainte, puisque tous les coûts sont positifs), soit proche de 0. Qui plus est, il a assez peu de sens économique. Le terme de droite de la deuxième inégalité est en revanche beaucoup plus contraignant. Cependant, il faut bien noter qu'en se plaçant dans un jeu à un seul tour, l'ensemble de l'investissement en capacité doit être amorti en une seule période avec un aléa conséquent. Si on modifie le modèle en considérant un jeu avec plusieurs étapes de confrontation d'offre et demande d'énergie (et donc plusieurs réalisations successives de la demande), il est possible de relâcher fortement cette condition sur les coûts. En effet, le profit d'une firme s'écrit alors (en notant t le nombre de périodes) :

$$E[\pi_i] = -I\bar{Q}_i + \sum_{l=1}^t \{(\bar{P} - C_p) \bar{Q}_i - (\bar{P} - C_p) \bar{Q}_i A\} = -I\bar{Q}_i + t \{(\bar{P} - C_p) \bar{Q}_i - (\bar{P} - C_p) \bar{Q}_i A\}$$

La condition nécessaire devient :

$$\frac{\partial E[\pi_i]}{\partial \bar{Q}_i} = -I + t \left\{ (\bar{P} - C_p) - (\bar{P} - C_p) A - (\bar{P} - C_p) \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} \right\} = 0$$

Soit :

$$1 - A - \bar{Q}_i \frac{\partial A}{\partial \bar{Q}_i} = \frac{I}{t(\bar{P} - C_p)}$$

La condition de deuxième ordre, quant à elle, est inchangée. L'intervalle de valeur sur lequel est définie la solution l'est donc également.

Dans cette version du modèle, les coûts d'investissement sont alors amortis sur plusieurs périodes, et la deuxième condition devient beaucoup moins contraignante :

$$\frac{I}{t(\bar{P} - C_p)} < 1 - P(D < 0)$$

Ce raisonnement est également valable pour le monopole public, ce qui permet de relâcher la contrainte sur ses coûts de telle sorte que :

$$\frac{I}{(C_D - C_p)} < t$$

En revanche, la première condition reste semblable (en terme de contrainte), puisque le terme de droite de l'inégalité est négatif.

$$\frac{I}{t(\bar{P} - C_p)} > 1 - \frac{M + \sqrt{M^2 + 16\sigma^2}}{4} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{-2M + \sqrt{M^2 + 16\sigma^2}}{4}\right)^2}{2\sigma^2}} - P(D < \frac{M + \sqrt{M^2 + 16\sigma^2}}{2})$$

Généralisation au cas d'un n-oligopole

Les calculs précédents ont été réalisés dans le cas d'un duopole afin de ne pas les rendre trop lourds, mais notons que ce résultat est valable si l'on suppose qu'à la place du duopole on a un n-oligopole. En effet, le programme de maximisation de la firme i est alors :

$$\max_{\bar{Q}_i} E[\pi_i] = \max_{\bar{Q}_i} E \left[(C_P \bar{Q}_i - C_P \bar{Q}_i - I \bar{Q}_i) 1_{\left\{ D < \sum_{j=1}^n \bar{Q}_j \right\}} + (\bar{P} \bar{Q}_i - C_P \bar{Q}_i - I \bar{Q}_i) 1_{\left\{ D \geq \sum_{j=1}^n \bar{Q}_j \right\}} \right]$$

La condition de premier ordre devient :

$$1 - \bar{Q}_i \frac{\partial P(D < \sum_{j=1}^n \bar{Q}_j)}{\partial \bar{Q}_i} - P(D < \sum_{j=1}^n \bar{Q}_j) = \frac{I}{(\bar{P} - C_P)}$$

L'équilibre est toujours symétrique, comme permet de le montrer la différence de la fonction de réaction de la firme i avec toutes les autres.

La condition de deuxième ordre devient :

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sum_{j=1}^n \bar{Q}_j - M)^2}{2\sigma^2}} \left(\bar{Q}_i^2 + \left(\sum_{j \neq i} \bar{Q}_j - M \right) \bar{Q}_i - 2\sigma^2 \right)$$

Si l'on se place à l'équilibre symétrique, on a : $\bar{Q}_i = \bar{Q}_j$, pour toutes les valeurs de i et j. En notant \bar{Q}^* la valeur à l'équilibre, le polynôme à considérer devient alors :

$$n\bar{Q}^{*2} - M\bar{Q}^* - 2\sigma^2$$

Son discriminant Δ est :

$$\Delta = M^2 + 8n\sigma^2$$

L'intervalle de valeurs pour lesquelles existe un maximum est très similaire au premier, puisqu'il comporte une racine positive et une racine négative :

$$r_1 = \frac{M + \sqrt{M^2 + 8n\sigma^2}}{2n} \quad r_2 = \frac{M - \sqrt{M^2 + 8n\sigma^2}}{2n}$$

Sur l'intervalle compris entre r_2 et r_1 la dérivée seconde est négative et la valeur d'équilibre pour une firme est un maximum. Dans le cas du duopole, cet intervalle a été considéré comme « raisonnable » dans la mesure où la valeur de r_1 relève de niveaux de demande extrêmement élevés. Dans le cas d'un n-oligopole, cet intervalle est légèrement moins contraignant, puisque avec l'augmentation du nombre d'entreprises, la proportion (en pourcentage) de la demande fournie par une entreprise décroît plus vite que la contrainte en r_1 . En effet, on peut grossièrement estimer la borne r_1 de telle sorte que l'inégalité devient de façon approximative :

$$n\bar{Q}^* < \frac{M}{2} + n\sqrt{\frac{2}{n}}\sigma^2$$

L'unicité de la solution se démontre de façon identique au cas du duopole. Il faut montrer que la fonction f suivante est monotone :

$$f(\bar{Q}_i) = 1 - \bar{Q}_i \frac{\partial P(D < \sum_{j=1}^n \bar{Q}_j)}{\partial \bar{Q}_i} - P(D < \sum_{j=1}^n \bar{Q}_j)$$

Or, sa dérivée est positive sur l'intervalle $[0, r_1]$ (elle a été calculée quelques lignes au-dessus). Donc, l'équilibre est unique sur cet intervalle.

La condition qui permet de comparer le niveau d'investissement du monopole public et celui de l'oligopole devient :

$$1 - P(D < n \bar{Q}_D^*) = 1 - P(D < \bar{Q}_M^*) + \bar{Q}_D^* \frac{\partial P(D < n \bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*)$$

Or, on a l'inégalité suivante :

$$\bar{Q}_D^* \frac{\partial P(D < n \bar{Q})}{\partial \bar{Q}_i}(\bar{Q}_D^*) > 0$$

Donc le niveau d'investissement de l'oligopole est inférieur à celui du monopole public. Par contre, comme on peut s'y attendre, plus le nombre d'entreprises augmente, plus le niveau d'investissement total augmente, et donc plus le niveau de capacité totale se rapproche du niveau du monopole public.

Généralisation à d'autres distributions de probabilité

Le modèle précédent peut être étendu, sous certaines conditions, à d'autres distributions de probabilité de la demande. En effet, exprimées en terme de probabilité, les conditions de premier ordre sont inchangées. Les principales modifications portent sur les conditions de deuxième ordre qui définissent l'intervalle des valeurs pour lesquelles l'équilibre est bien optimal. Le cas du monopole public est assez simple : la seule condition est que la dérivée de la probabilité (que la demande soit supérieure aux capacités installées) soit négative. Dans le cas du duopole privé, il n'est pas possible de se donner une condition de validité. En effet, la condition de deuxième ordre s'écrit :

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = -\bar{Q}_i \frac{\partial^2 A}{\partial \bar{Q}_i^2} - 2 \frac{\partial A}{\partial \bar{Q}_i}$$

Or cette équation n'est pas de signe constant, même dans le cas de la loi normale. En revanche, le résultat selon lequel l'investissement du monopole public sera toujours supérieur à celui du duopole, quel que soit le niveau du prix plafond, tient toujours, puisqu'on a :

$\bar{Q}_D^* \frac{\partial P(D < 2\bar{Q}_D^*)}{\partial \bar{Q}_i}(\bar{Q}_D^*) > 0$, ce qui est vérifié pour n'importe quelle densité de probabilité.

c-Conclusions

En comparant les valeurs d'investissement d'équilibre pour le monopole public et pour le duopole déréglementé, il est possible de conclure que le niveau d'investissement sera systématiquement inférieur dans le cas du duopole. Mieux ce résultat est aussi valable dans le cas d'un n-oligopole. Ce modèle d'investissement, bien que très simplifié, a néanmoins le mérite de mettre l'accent sur deux aspects non examinés dans la littérature sur les marchés d'électricité : premièrement, l'aléa très important qui affecte les revenus des investissements ; deuxièmement, les problèmes de capacité. Ces derniers engendrent l'attrait important qu'il peut y avoir à sous investir afin d'accroître la probabilité des pics de prix, alors même que le niveau de ces derniers est fixé de façon exogène.

Le terme avec dérivée de la probabilité qui s'ajoute à l'expression du niveau de capacité optimal du duopole est assimilable à la prise en compte de l'externalité positive pour la firme que constitue la hausse de son profit espéré généré par une baisse de sa capacité de production. L'explication principale de ce résultat tient dans la différence fondamentale entre monopole et firme privée : l'un n'a que des coûts à prendre en compte dans son optimisation, tandis que l'autre voit une partie de ce qui constitue un coût pour le monopole (la demande non fournie au coût marginal) lui revenir sous forme de profit. Une réduction des montants d'investissement induit directement pour le monopole une hausse des coûts espérés de fournitures. Cette même baisse pour une firme a deux effets contraires : d'une part, elle réduit les gains futurs en diminuant sa part de marché en cas de sous-capacité ; d'autre part, elle augmente la probabilité qu'il y ait un phénomène de sous capacité et donc augmente potentiellement aussi ses gains futurs. Nous avons vu que le deuxième effet empêche qu'il y ait la même capacité de pointe totale dans le cas d'un duopole déréglementé que dans le cas d'un monopole public. Et cela, même si le niveau de prix en cas de sous capacité est égal au coût de

défaillance (ce n'est qu'un cas limite évidemment puisque les consommateurs sont alors indifférents entre consommer et ne pas consommer).

Le résultat tient-il dans l'absence d'élasticité de la demande pour le monopole ? Non, car fixer le prix exogène au coût de défaillance est équivalent à une absence d'élasticité pour le duopole et le résultat est toujours valable. Par ailleurs, il faut tenir compte du fait que les tarifs particuliers qui permettent à un monopole public d'obtenir une élasticité de la consommation, augmentent ses coûts fixes, ce qui n'est pas vrai pour le duopole déréglementé.

La comparaison des deux modèles met en évidence que le pouvoir de marché détenu par les entreprises à travers le processus d'investissement conduit à un investissement sous optimal. Par rapport à l'approche en terme de fonctions de profit, où le « VOLL pricing » (le prix plafond est fixé à la valeur du coût de défaillance et il n'y a pas de politique de gestion sur prix des réserves opérationnelles) permettait d'induire un niveau d'investissement optimal, l'introduction d'un pouvoir de marché vient invalider ces conclusions. Ce résultat est peu étonnant, surtout au regard de la structure oligopolistique du marché. Cependant, comme nous l'avons déjà souligné, une telle structure de marché à la pointe est tout à fait plausible, surtout si l'on tient compte des délais de construction de nouvelles installations (3 ans en général). Ces derniers légitiment tout à fait cette situation, même dans un marché où existent des incitations suffisantes à l'investissement. Car, d'une part, un certain laps de temps est nécessaire avant que les incitations ne produisent leurs effets ; d'autre part, parce qu'il peut très bien ne pas exister suffisamment de demande pour plus de producteurs à la pointe. De plus, en situation de tension entre l'offre et la demande, plus on se rapproche de la pointe, plus le nombre de producteurs pouvant fournir les derniers mégawatts de puissance a tendance à diminuer. La pointe est donc particulièrement sensible à la présence d'un pouvoir de marché par l'investissement. Un tel pouvoir de marché est de plus extrêmement difficile à repérer. Il suppose en effet de connaître particulièrement bien les lois de

probabilités de la demande, ce qui est difficile eu égard aux problèmes d'estimation et d'inférence statistique. Mais ce constat vaut aussi pour les producteurs. Autrement dit, le sous-investissement peut très bien résulter d'une aversion au risque des producteurs qui ne sont pas à même de correctement estimer leur espérance de rente de rareté.

Le modèle indique par ailleurs clairement que certaines solutions invoquées pour résoudre le problème des « revenus manquants » sont inefficaces. Ainsi, diminuer le coût d'investissement pour une firme privée par le biais de subventions va augmenter l'investissement, mais génère un coût global pour la collectivité (subventions plus investissement privé) plus important que si le même investissement en capacité était effectué par un monopole public. De même, le coût social de la défaillance (ou plutôt de l'augmentation de la probabilité qu'il y ait défaillance) ne peut être intégré par l'intermédiaire d'un prix qui serait imposé à l'oligopole. Ce résultat est assez fort, puisqu'il implique qu'il est impossible de parvenir à un optimum de second rang par le seul jeu de prix réglementés. Ainsi, tout pouvoir de marché par l'investissement (et plus particulièrement à la pointe) empêche qu'un mécanisme de marché restaure les incitations à l'investissement et permette d'obtenir la capacité d'investissement optimale. Que cette capacité ne soit pas atteinte ne signifie pas nécessairement la survenue de délestages, mais implique que la probabilité de leur occurrence augmente. Autrement dit, engendre une dégradation de l'aptitude du système électrique considéré à fournir la demande.

Enfin, le modèle présenté dans les pages précédentes montre clairement les limites de l'approche en terme de fonction de profit. Bien que permettant une simplification utile du processus de fixation des prix, celle-ci fait complètement l'impasse sur le pouvoir de marché par l'investissement. Or, ce pouvoir constitue un phénomène primordial à intégrer dans toute réflexion sur les « revenus

manquants », puisqu'il est à même de perturber le fonctionnement des dispositifs conçus pour y remédier, comme nous allons le voir de façon plus détaillée.

Partie 3 : Les réserves opérationnelles

Les réserves opérationnelles constituent une caractéristique particulière des systèmes électriques qui jouent un rôle très important dans la stabilité de ces derniers. En effet, une partie importante des capacités d'un système électrique doit être en permanence mobilisée pour pouvoir répondre en temps réel aux contingences affectant ce système (Joskow et Tirole (2006) citent le chiffre de 10-12 % des capacités installées, auxquelles il faut ajouter 5-6 % d'indisponibilités en moyenne pour obtenir le chiffre souvent cité de 118 % de la pointe comme estimation des capacités à déployer). Les réserves sont donc des unités de production classiques, sélectionnées pour être utilisées comme réserves opérationnelles par le gestionnaire de réseau.

Dans la mesure où il est impossible de déterminer que l'énergie produite en un point du réseau ira en un autre point précis et dans la mesure où offre et demande doivent en permanence être égales, le dispatching (planification) des moyens de production se fait la veille pour le lendemain. Cependant, comme la demande n'est vraiment connue qu'en temps réel, et comme des avaries imprévues peuvent affecter les moyens de production, il est nécessaire de disposer de moyens de production à même de réagir rapidement afin de maintenir l'équilibre offre-demande à chaque instant. C'est le rôle des réserves opérationnelles, dont une caractéristique importante consiste justement dans la rapidité à être mobilisées. Il n'est donc pas possible d'employer pour cela une centrale nucléaire, son démarrage et / ou son arrêt étant gourmands en temps, sans même mentionner le fait qu'une centrale nucléaire, une fois chargée en combustible est censée produire jusqu'à épuisement de celui-ci (on parle de « campagne » de production). Les différents types de centrales thermiques (fuel, charbon gaz) ont des potentialités plus ou moins importantes pour pouvoir avoir le rôle de réserve opérationnelle ; les plus adaptés étant les moyens de production hydrauliques, de grande réactivité (et

justement ceux qu'on ne peut construire à volonté ...). Les moyens utilisés comme réserves sont donc fréquemment des unités de pointe, dont on a vu qu'elles étaient les plus difficiles à rentabiliser. Les marchés censés permettre au gestionnaire de réseau d'acheter des réserves opérationnelles classent d'ailleurs les différents moyens de production éligibles suivant leurs vitesses de réaction (par exemple, les moyens mobilisables en moins de dix minutes, ceux mobilisables en moins de trente minutes), chaque catégorie représentant un segment des marchés.

De plus, le fait de ne pouvoir disposer de réserves suffisantes peut se traduire par un effondrement du réseau. Economiquement, cela signifie que plus aucune unité de production ne peut produire et qu'aucun consommateur ne peut être servi. En outre, rétablir le fonctionnement normal du réseau prend beaucoup de temps. La nécessité de disposer de réserves opérationnelles impose donc de fortes contraintes à la gestion d'un système électrique, aussi bien techniques qu'économiques.

Dans la mesure où les réserves sont des moyens de production essentiels pour assurer l'équilibre offre-demande, elles font partie à part entière de la problématique qui nous occupe. Nous nous proposons donc d'essayer de les intégrer à notre réflexion en trois temps : d'abord, en nous intéressant à la possibilité évoquée par Stoft (2002) de les utiliser pour modifier la durée des pics de prix, ce qui nous amène aux conséquences de leur statut de bien public évoqué par certains auteurs (Joskow, Tirole 2006 ; Meunier, Finon 2006) ; puis en intégrant ces conséquences dans une réflexion sur les récents designs de marchés de capacité, explicitement conçus pour remédier au problème des « revenus manquants » ; enfin, en étudiant la possibilité que les moyens de pointe soient détenus et opérés par le gestionnaire de réseau.

a-Le rôle fondamental des réserves opérationnelles

i. Une politique de gestion des pics de prix est-elle possible ?

Il est intéressant de revenir plus en détail sur le raisonnement de Stoft à partir duquel il construit la fonction du profit espérée d'un investissement en pointe. Ce raisonnement est présenté dans son ouvrage « Power System Economics » (2002), mais de façon un peu plus détaillée dans un article de 2003.

Stoft part de deux constats : le premier est qu'il n'est pas rationnel dans un système déréglementé de chercher à tout prix à maintenir les réserves opérationnelles au niveau exigé par les normes techniques. Celles-ci doivent avoir un coût, ou tout au moins il doit exister un niveau de prix au-delà duquel il n'est pas rationnel pour le gestionnaire de réseau d'acheter des réserves. Soit un prix plafond pour l'achat de ces dernières. Le deuxième constat est que les producteurs ont le choix entre produire et vendre de l'énergie sur le marché spot, ou bien vendre de la réserve. Cette dernière option permet d'économiser le coût marginal du système. Mais du fait du choix entre les deux opérations, il est possible d'arbitrer entre les deux marchés. Il en conclut à l'existence d'une relation d'égalité (approximative, étant donné que toute unité de production n'est pas éligible comme réserve) entre le prix sur le marché spot et le prix plafond des réserves auquel est ajouté le coût marginal du système. Ainsi, le prix plafond des réserves plafonne également le prix du marché spot. Ce point est important : la politique de gestion des réserves opérationnelles a un effet important sur le marché spot. Il n'est pas possible de les considérer uniquement sous leur aspect technique, puisqu'elles ont un impact économique fort.

Pour une unité de pointe, placée en haut du merit order, le coût variable peut être considéré comme égal au coût marginal du système. Or, le profit dégagé de la vente d'une unité d'énergie par une centrale de pointe va souvent être égal au prix

plafond du marché moins son coût variable, c'est-à-dire au prix plafond des réserves. Stoft suppose de façon implicite qu'en dehors des périodes de tension sur le marché spot (où le prix atteint son plafond), une centrale de pointe vendra à son coût variable. Cette hypothèse avait déjà été soulignée dans la partie consacrée à l'approche en terme de fonction de profit. Stoft avance que l'approximation entre profit de la vente d'une unité d'énergie et prix des réserves est de l'ordre de quelques dollars par mégawatheure, quand l'ordre de grandeur des prix plafond est de 1000 dollars par mégawatheure.

La construction de la fonction de profit se fait ensuite en posant qu'autour de la pointe, il est possible de retenir la relation suivante entre demande « augmentée », c'est-à-dire incluant les indisponibilités des centrales comme une hausse de la demande (en puissance) et durée d'un niveau de puissance donnée :

$$L_g(D) = L_{\max} - \alpha D$$

Si l'on suppose maintenant que le prix plafond sur le marché spot n'est atteint que lorsqu'il existe un déficit de réserves opérationnelles (entraînant des achats de puissance de la part du gestionnaire de réseau), on peut obtenir la durée des pics de prix sur ce marché par l'égalité entre les capacités installées (K) et la demande, majorée du niveau de réserves opérationnelles exigée (RO^E) :

$$L_g(D_{PP}) + RO^E = K$$

La durée des pics de prix est donnée par la substitution de la première équation dans la deuxième. On obtient alors :

$$D_{PP} = \frac{1}{\alpha} (L_{\max} + RO^E - K)$$

Et finalement, la fonction de profit d'une unité de pointe (exprimée en €/MWh) est donnée par le produit de cette durée par le prix plafond (\bar{P}) :

$$\pi = \bar{P} \cdot D_{PP}(K) = \bar{P} \frac{1}{\alpha} (L_{\max} + RO^E - K)$$

On obtient donc une relation entre les capacités installées et le profit annuel d'une unité de pointe. L'hypothèse centrale de Stoft est que le niveau d'investissement sera donné par l'égalité entre cette fonction et les coûts fixes annuels de l'unité (exprimés en €/MWh, ce qui suppose d'utiliser une durée de fonctionnement pour les obtenir). Ces derniers incluant évidemment le taux d'actualisation qui a permis de les calculer et qui est donc égal au taux de rémunération du capital exigé par les investisseurs.

Autrement dit, dès qu'il est possible de réaliser un surprofit ($\pi > \text{Coûts Fixes}$), un ou des investisseurs installent les capacités nécessaires, diminuant la durée des pics de prix et annulant le surprofit. Or, autant une telle relation d'arbitrage peut se justifier sur un marché financier, autant elle semble irréaliste sur un marché électrique, en particulier à la pointe (comme nous l'avons montré dans la partie précédente). Néanmoins, cette relation détermine le niveau d'investissement en fonction des deux paramètres politiques que sont le prix plafond et le niveau exigé des réserves opérationnelles. Il est donc possible, en jouant sur ces deux paramètres, d'obtenir le niveau d'investissement souhaité (qui doit être préalablement déterminé).

Le raisonnement pour justifier une politique de gestion des pics de prix est justement qu'un prix plafond bas (nécessaire pour compenser les phénomènes de pouvoir de marché), qui crée le problème des « revenus manquants », peut être compensé par l'élévation du seuil des réserves opérationnelles (qui déclenche le passage du prix spot au prix plafond). Ce raisonnement tient-il toujours s'il existe un oligopole pour la fourniture de la pointe ? Un tel oligopole serait de nature à détenir un pouvoir de marché d'investissement (comme dans le modèle de la partie précédente).

Supposons donc que la forme de la fonction de profit trouvée précédemment s'applique toujours, mais s'applique à la firme i d'un n -oligopole de Cournot. Pour que les résultats s'interprètent plus intuitivement, on va noter \bar{P} le prix plafond du marché spot (et non celui des réserves). On obtient alors :

$$\pi_i = (\bar{P} - C_P) \cdot D_{PP}(K_i + \sum_{j \neq i} K_j) = (\bar{P} - C_P) \frac{1}{\alpha} (L_{\max} + RO^E - K_i - \sum_{j \neq i} K_j)$$

Comme cette fonction renvoie un résultat en €/MWh, et que l'on souhaite intégrer le coût de l'investissement (en supposant qu'il est de I par unité de capacité installée, comme dans le modèle de la partie précédente), il est nécessaire de multiplier cette écriture par K_i . On obtient donc la fonction de profit total de la firme i :

$$\pi_i = (\bar{P} - C_P) \frac{1}{\alpha} (L_{\max} + RO^E - K_i - \sum_{j \neq i} K_j) K_i - IK_i$$

D'où la condition de premier ordre :

$$\frac{\partial \pi_i}{\partial K_i} = (\bar{P} - C_P) \frac{1}{\alpha} (L_{\max} + RO^E - 2K_i - \sum_{j \neq i} K_j) - I = 0$$

Cette égalité définit implicitement la fonction de réaction de la firme i . La condition de deuxième ordre est alors :

$$\frac{\partial^2 \pi_i}{\partial K_i^2} = -2(\bar{P} - C_P) \frac{1}{\alpha} < 0$$

La condition de premier ordre nous donne bien un maximum. En opérant la soustraction des fonctions de réaction des différentes firmes, on montre que l'équilibre est symétrique. La condition de premier ordre s'écrit :

$$(\bar{P} - C_p) \frac{1}{\alpha} (L_{\max} + RO^E - (n+1)K_i) - I = 0$$

la valeur d'équilibre de la capacité de la firme i étant :

$$K_i^* = \frac{1}{(n+1)} (L_{\max} + RO^E - \frac{\alpha I}{(\bar{P} - C_p)})$$

Le montant total des capacités installées est alors :

$$K_{priv}^* = \frac{n}{(n+1)} (L_{\max} + RO^E - \frac{\alpha I}{(\bar{P} - C_p)})$$

Cette valeur est à comparer avec le niveau d'investissement d'un monopole public dans ce modèle (remplacer le prix plafond par la valeur du coût de défaillance permet de mieux saisir la logique de cette transposition), qui est déterminé par l'égalité :

$$(\bar{P} - C_p) \cdot D_{pp}(K) = (\bar{P} - C_p) \frac{1}{\alpha} (L_{\max} + RO^E - K) = I$$

Soit, un niveau d'investissement optimal :

$$K_{pub}^* = L_{\max} + RO^E - \frac{\alpha I}{(\bar{P} - C_p)}$$

La comparaison des deux valeurs montre immédiatement qu'à prix plafond et niveau des réserves opérationnelles identiques, l'oligopole investira moins qu'un monopole public. Ces résultats sont identiques à ceux de la partie précédente. Comme le niveau d'investissement d'un monopole public n'a de sens que si le prix plafond est au niveau du coût de la défaillance, c'est cette valeur qui est retenue pour définir le niveau d'investissement optimal. Notons en revanche que le niveau des réserves opérationnelles dans l'équation qui donne l'investissement optimal du monopole a une valeur fixe qu'on va noter RO_M^E . Dès lors, les valeurs des paramètres politiques qui permettent d'atteindre ce niveau dans un marché déréglementé vont être données par l'égalité suivante :

$$L_{\max} + RO_M^E - \frac{\alpha I}{(C_D - C_P)} = \frac{n}{(n+1)} (L_{\max} + RO^E - \frac{\alpha I}{(\bar{P} - C_P)})$$

Il y a deux paramètres à déterminer et une équation. Supposons donc que le prix plafond est fixé par ailleurs par des considérations de pouvoir de marché (au sens traditionnel) à un niveau largement inférieur au coût de la défaillance. L'équation précédente nous permet alors d'obtenir le niveau de réserves opérationnelles nécessaire. On obtient :

$$-\frac{n}{(n+1)} RO^E = \frac{n}{(n+1)} (L_{\max} - \frac{\alpha I}{(\bar{P} - C_P)}) + \frac{\alpha I}{(C_D - C_P)} - L_{\max} + RO_M^E$$

Soit :

$$RO^E = \frac{n+1}{n} (L_{\max} + RO_M^E - \frac{\alpha I}{(C_D - C_P)}) - (L_{\max} - \frac{\alpha I}{(\bar{P} - C_P)})$$

Il semble donc possible d'atteindre le niveau d'investissement optimal en augmentant le seuil au-delà duquel est déclaré un déficit des réserves opérationnelles. Les incitations à investir sont restaurées. Notons que la mise en place de cette mesure revient à supprimer la fixation des prix par le marché, et ce

sur une partie d'autant plus importante de la monotone de production que la structure de fourniture de la pointe est oligopolistique. Elle s'apparente à bien des égards à une gestion administrative de la demande à partir d'un certain seuil.

Joskow (2006b) semble aller dans ce sens quand il propose d'élever le prix spot (et celui des réserves, avec les coûts supplémentaires pris en compte) au niveau du prix plafond dès qu'il existe un déficit des réserves opérationnelles. Le modèle précédent illustre cette hypothèse. Cela constitue de plus un moyen d'intégrer certains éléments concernant la gestion des réserves opérationnelles. En effet, comme Joskow le rappelle, la première mesure prise avant un délestage dans une situation de déficit des réserves est la réduction du voltage sur le réseau de 5 %. Or, cette décision, qui affecte le fonctionnement des appareils électriques, réduit la demande et donc les prix sur le marché spot. Il est donc presque impossible de refléter le coût de cette mesure dans les prix, si ceux-ci sont déterminés par le marché. Imposer que le prix spot soit au niveau du prix plafond dès qu'un déficit est constaté est une mesure simple qui permet de le faire partiellement. Cependant Joskow ne recommande à aucun moment de jouer sur le seuil des réserves pour élever la durée des pics de prix (de même que Stoft n'est pas revenu sur ses propositions dans ses articles postérieurs).

En effet, corriger le problème des « revenus manquants » par cette mesure pose plusieurs problèmes. Tout d'abord, il est nécessaire d'estimer correctement quel est le pouvoir oligopolistique à la pointe (le nombre d'entreprises dans le modèle). Celui-ci peut changer suivant les niveaux de demande. Or, estimer correctement quel est le seuil à fixer dépend crucialement de ce paramètre. De plus, cette politique risque de vider de substance le fonctionnement du marché pour une part importante de la courbe de charge, sans pour autant remédier au pouvoir de marché par l'investissement, et cela d'autant plus que le prix plafond sera bas relativement au coût de la défaillance. Ensuite, cela risque d'entraîner une surévaluation des réserves. En effet, soit le seuil est élevé sans qu'il y ait d'achats de la part du gestionnaire de réseau et la plupart des générateurs pouvant fournir des réserves

vont s'orienter sur le marché spot, soit il y a des achats inutiles de réserves, ce qui conduit à sur-rémunérer ces générateurs. Autant imposer le prix plafond sur le marché spot ne pose pas de problème du fait qu'il y aura toujours des achats de la part des distributeurs, autant sur le marché des réserves cela est problématique puisqu'il n'y a qu'un seul acheteur, le gestionnaire de réseau. Enfin, une telle politique comporte un risque important en terme de crédibilité vis-à-vis des investisseurs, puisque ces paramètres doivent constamment être révisés pour induire le bon niveau d'investissement. Elle est donc fortement sujette à des changements d'opinions du régulateur et du gestionnaire de réseau. De fait, elle ne peut pas constituer un moyen efficace pour restaurer à long terme les incitations à investir.

Tous ces éléments ne sont pas capturés dans le modèle dans la mesure où les réserves opérationnelles n'y ont pas d'autre rôle que de permettre d'élever artificiellement la demande. Autant la réflexion de Stoft est stimulante et permet de mettre en évidence certaines conséquences notables de la gestion des réserves, autant ses conclusions de 2002 et 2003 ne semblent pas pouvoir fournir une réponse satisfaisante au problème des « revenus manquants ». Ceci explique sans doute pourquoi il s'est focalisé par la suite sur la résolution de ce problème par le biais des marchés de capacité. Cependant, avant de nous intéresser à ces derniers, il est utile de poursuivre l'analyse économique des réserves opérationnelles.

ii. Les réserves opérationnelles sont un bien public

Une caractéristique des réserves opérationnelles qui n'a pas encore été évoquée est lourde de conséquences économiques : il s'agit de leur rôle pour éviter un effondrement du réseau. En effet, le risque de disposer de capacités de production insuffisantes ne se résume pas seulement au fait d'être obligé de délester une partie de la demande. Il implique également une hausse de la probabilité que le gestionnaire de réseau ne soit pas capable de faire face à des variations trop importantes de la demande (ou de l'offre), de telle sorte que le réseau électrique s'effondre. De tels effondrements ont pu faire l'actualité sur les trois dernières années (Italie 2003, Nord-Est des Etats-Unis en 2003 également, mais aussi en Europe du Nord en 2006), sans que tous ces cas soient imputables à un manque de capacités de production (les problèmes d'investissement en capacités de distribution ou de mauvaise gestion d'un réseau sont également des causes d'effondrement). Néanmoins, l'occurrence de ces événements dans un contexte où le sous-investissement en moyens de production commence à être reconnu permet utilement d'en rappeler l'importance.

C'est que l'effondrement du réseau peut affecter l'intégralité d'un système électrique. Dans ce cas, ce n'est plus seulement une partie de la demande qui ne peut être servie, mais l'intégralité. Le coût collectif est alors particulièrement élevé (l'intégralité de la demande multipliée par le coût de la défaillance). De plus, les producteurs qui pouvaient être mobilisés ne peuvent produire, alors même qu'un tel événement a de fortes chances de survenir pendant les heures de l'année censées rémunérer une bonne partie des coûts fixes d'une unité de pointe par le biais des rentes de rareté. Lors d'un délestage un surcroît de capacité a une forte valeur, tandis que lors d'un effondrement, sa valeur est nulle. Comme le remarquent Joskow et Tirole (2006), il y a une externalité imposée aux autres générateurs par celui ou ceux qui sont à l'origine de la situation. Enfin, restaurer le fonctionnement normal du réseau prend énormément de temps, ce qui accroît très

fortement les coûts d'un tel effondrement par rapports aux délestages tournants dont le coût social est déjà considéré comme très élevé (par le biais du coût de la défaillance). Il y a là un phénomène qui ne peut pas ne pas être inclus dans la réflexion sur le processus d'investissement dans le secteur électrique.

Mieux, Joskow et Tirole (2006) expliquent qu'en raison de l'importance des coûts économiques d'un effondrement, et de la difficulté à intégrer ces coûts par le biais d'un mécanisme de marché, les réserves opérationnelles ont les attributs d'un bien public. Meunier et Finon (2006) reconnaissent également cette caractéristique des capacités de pointe, aussi bien pour leur rôle comme réserves opérationnelles que pour leur contribution à éviter des délestages. En effet, les incitations à investir provenant des pics de prix n'incluent pas le gain social attendu d'une meilleure capacité à éviter des délestages et / ou un effondrement, alors même que tout consommateur bénéficie des décisions privées d'investissement. De plus, comme le notent Joskow et Tirole (2006), la possibilité d'un effondrement rend le problème plus complexe dans la mesure où les usagers du réseau considèrent sa stabilité comme exogène à leurs décisions. C'est la deuxième raison qui plaide pour considérer les réserves opérationnelles comme un bien public. Le gain en stabilité apporté par un surcroît de capacités éligibles comme réserve profite à tous les utilisateurs du réseau électrique considéré, tandis que les coûts d'investissement en sont supportés par un seul investisseur. Il s'agit donc bien d'une situation typique définissant un bien public. Cependant, dans la mesure où les réserves sont des moyens de production « classiques » qui peuvent aussi bien être utilisés sur le marché de l'énergie, il est a priori plus difficile de recourir aux moyens traditionnels de prise en compte des biens publics (propriété publique, financement par impôt).

Joskow et Tirole (2006) montrent qu'il est possible de déterminer un optimum de Ramsey tenant compte de la spécificité des réserves opérationnelles. La partie 4 de leur article qui traite de cet aspect est fondamentale à bien des égards. Il s'agit, en effet, de la première analyse théorique du rôle des réserves opérationnelles. Nous

allons donc brièvement revenir sur leurs résultats, afin de mieux cerner l'intérêt de leur solution.

Le modèle que proposent Joskow et Tirole (2006) relève d'une logique assez similaire à celui présenté dans le deuxième chapitre. Cependant, les réserves opérationnelles y ont un rôle important à travers une représentation très simple du dispatching et seul l'optimum de Ramsey est décrit. Il n'existe qu'un seul type de technologie de générateur, dont le coût de construction de chaque unité de capacité est constant, et dont le coût variable est donné. La demande est aléatoire et complètement inélastique, mais sa loi de probabilité n'est pas précisée. Le modèle comporte une autre variable aléatoire représentant la disponibilité de la capacité. Cette variable a une distribution telle que la probabilité que la disponibilité soit égale à 100 % est non-nulle (alors, que l'intervalle des valeurs de la variable est continu). De plus, cette variable va être considérée comme indépendante des états de la nature (donc de la demande), ce qui est une simplification importante, puisqu'il y a de fortes chances qu'un effondrement survienne lorsque la demande est élevée.

La séquence des décisions dans le modèle est la suivante. L'investissement a lieu préalablement à la réalisation de la demande et de la disponibilité. Puis, la demande est connue et le gestionnaire de réseau peut choisir de délester une partie de la demande (ce qui a un coût social par unité d'énergie égal au coût de la défaillance) et d'affecter une partie des capacités aux réserves opérationnelles en choisissant le taux de réserve (le fait de dispatcher une unité de capacité a un coût). La probabilité d'un effondrement du système est alors donnée par la probabilité que le dispatching soit impossible du fait de l'indisponibilité, donc par la probabilité que le produit de la disponibilité et du taux de réserve (auquel on ajoute un) soit inférieure à un. Enfin, la disponibilité est connue : soit le système s'effondre et l'intégralité de la demande n'est pas servie, soit il ne s'effondre pas et le dispatching est réalisé.

Sans connaître la distribution de la demande (dont les états de la nature sont indicés par i), il est possible de résoudre le programme d'un planificateur

maximisant le surplus social en jouant sur plusieurs variables de décision : le niveau de capacité, la part de la demande délestée dans l'état de la nature i et le taux de réserve dans le même état de la nature.

Cette maximisation permet de définir trois régimes de fonctionnement en fonction des valeurs possibles de la demande. Quand cette dernière est suffisamment faible pour qu'une part importante des capacités servent de réserves, aucun délestage n'est opéré et le taux de réserve prend une valeur constante sur toute la zone, correspondant au niveau maximum des réserves. La probabilité d'un effondrement est minimale. Si la demande s'élève de telle sorte qu'il ne soit plus possible de maintenir ce niveau des réserves, elle continue à être complètement servie et le taux de réserve diminue jusqu'à un minimum qui permet de définir le passage dans la troisième zone. Plus la demande s'élève dans cette zone, plus la probabilité d'un effondrement augmente. Cette dernière est définie par une demande trop élevée par rapport aux capacités pour qu'il n'y ait pas de délestage, de telle sorte que le taux de réserve prend une valeur constante sur toute la zone et que toute hausse de la demande se traduit par une augmentation équivalente du délestage. La probabilité d'un effondrement est constante sur toute la zone et maximale.

Le programme d'investissement optimal nous donne les prix qui permettraient d'implémenter l'optimum de Ramsey. Dans la première zone, le prix doit être égal au coût variable des capacités. Dans la troisième zone, il doit être égal au coût de la défaillance, puisqu'une augmentation des capacités d'une unité permet de réduire la demande non servie, ce qui augmente le surplus social du coût de la défaillance. Enfin, la zone intermédiaire connaît un prix croissant avec la demande, évoluant du coût variable au coût de la défaillance. Ce prix représente bien évidemment l'augmentation de surplus social induite par l'augmentation d'une unité de capacité en terme de réduction de la probabilité d'un effondrement du système et en terme de réduction de la probabilité de délestage. De tels prix permettent de couvrir les coûts fixes des capacités.

Néanmoins, l'implémentation du prix dans la zone intermédiaire (de réduction des réserves opérationnelles) est assez délicate. En effet, dans cette zone, la courbe de demande et la courbe d'offre sont verticales et confondues. La moindre erreur dans l'évaluation des réserves entraîne dès lors des variations très importantes du prix. Joskow et Tirole (2006) appellent ce problème « knife edge ».

Ce modèle, qui représente à bien des égards une avancée particulièrement intéressante, ne permet cependant pas d'obtenir le comportement d'investissement d'une firme, même en spécifiant les lois de probabilité des deux aléas. Cependant, quelques remarques peuvent être faites sur la solution permettant l'implémentation de l'optimum de Ramsey. Nous retrouvons ici les conclusions négatives de la deuxième partie : l'existence d'un pouvoir de marché par l'investissement empêche l'implémentation de l'optimum social par le biais du prix de l'énergie. En effet, les entreprises opérant avec ce prix ne peuvent réaliser de profit que si elles se situent dans les deux dernières zones. Dans la zone de délestage, l'opération est même particulièrement intéressante puisqu'elles seront payées au coût de la défaillance, avec un risque d'effondrement du système non négligeable néanmoins. Dans la mesure où c'est le niveau d'investissement qu'elles choisiront qui déterminera la probabilité d'occurrence des différents régimes, elles vont avoir tendance à essayer d'en maximiser la probabilité, ce qui signifie restreindre leurs capacités. Mieux, les prix imposés sont tels qu'une entreprise ne peut réaliser de profit qu'à partir de la zone où la probabilité d'effondrement augmente. Leur choix d'investissement va donc certainement conduire à augmenter celle-ci par rapport au niveau donné par l'optimum de Ramsey. Son implémentation pratique semble donc problématique, ce qui laisse ouverte la question du traitement des réserves opérationnelles.

La non prise en compte du coût social d'une insuffisance d'investissement s'avère encore plus problématique si l'on intègre cette nouvelle donnée. Cet aspect plaide très nettement pour une gestion des réserves opérationnelles par les pouvoirs publics. Cependant la forme exacte que doit prendre cette intervention est difficile

à établir sans analyse théorique : faut-il subventionner les moyens de productions éligibles comme réserves, ou bien faut-il que le gestionnaire de réseau dispose en propre de moyens de production dédiés à une utilisation comme réserves (comme le proposent implicitement Meunier et Finon (2006)) ?

b-Les nouveaux designs de marché de capacité

Les marchés de capacités sont des institutions mises en place afin de faire en sorte que le niveau des capacités totales sur un marché donné atteigne le niveau considéré comme optimal. Ils ont été introduits dans les Etats du Nord-est des Etats-Unis suite à la déréglementation de ces systèmes électriques. Leur intérêt est de reconnaître la nécessité de rémunérer la capacité indépendamment des pics de prix sur le marché de l'énergie. A ce titre, ils peuvent constituer une réponse adéquate aux problèmes de « revenus manquants », spécialement à la lumière des problèmes qui ont été soulevés précédemment.

i. Les designs « convergents »

Tout marché de capacité repose sur la fixation par les autorités de régulation d'un objectif de capacités à atteindre tenant compte des besoins de capacité nécessaires au bon fonctionnement d'un système électrique. Une valeur traditionnelle pour un tel niveau est 118 % de la valeur de la pointe attendue, valeur supérieure à la pointe afin d'inclure les indisponibilités et les besoins en réserves opérationnelles. Dans les marchés de capacité « traditionnels » (souvent appelés ICAP), les principaux mécanismes sont les suivants. Le niveau optimal est fixé par le régulateur, et il induit la demande en capacité sur le marché par le biais des objectifs de capacité individuels imposés aux différentes entreprises qui fournissent les consommateurs (ou plus généralement aux « Load Serving Entities » pour reprendre la terminologie américaine). Ces objectifs peuvent être atteints soit en disposant de capacités de production, soit en passant des contrats pour l'utilisation de capacités de tiers. En cas de non respect de ces obligations, les entreprises doivent payer une pénalité proportionnelle à la capacité manquante (et donc exprimée en €/MW). A titre d'exemple, cette pénalité a été fixée à 177 USD/MWjour sur le PJM, un niveau sensé refléter les coûts fixes rapportés à une journée d'une unité de pointe.

Il existe une deuxième pénalité si les entreprises ne sont pas à même de produire lorsqu'elles sont appelées. Le marché de capacité rend donc possible l'échange des capacités à un prix variable suivant l'évolution des besoins.

Une telle structure de marché pose deux problèmes : le premier est le caractère heurté suivi par les prix. En effet, lorsque les capacités disponibles sont supérieures au niveau requis par le régulateur, en l'absence de pouvoir de marché, le prix de la capacité tend à être nul. Tandis que lorsque les capacités sont inférieures au niveau requis, le prix est égal à la pénalité, puisque c'est le montant de la perte évitée à une entreprise par la capacité supplémentaire. Hors pouvoir de marché, les prix sur un marché de capacité tendent donc à connaître des sauts entre zéro et la valeur de la pénalité. Cela pose problème dans la mesure où un tel saut dans la rémunération potentielle d'un investisseur crée le risque pour ce dernier d'annuler celle-ci dans le cas où d'autres projets mis en œuvre en même temps permettent de dépasser le niveau objectif de capacité. Ce problème est reconnu dans le « Capacity Markets White Paper » de la CPUC (2005), où les rédacteurs indiquent clairement (p 20) que :

« The slope of fixed-cost recovery curve should be set so there are no large differentials in fixed-cost payments between slight shortages and slight surpluses. »

Ils suivent en cela les réflexions de Cramton et Stoft (2006) sur le design d'un marché de capacité, qui recommandent d'établir explicitement la fonction de profit d'un investisseur potentiel, fonction qui est appelée « fixed-cost recovery curve » dans le « Capacity Markets White Paper ». Une telle variabilité des prix résulte de la forme verticale implicite qu'a la courbe de demande sur ces marchés. De plus, le saut potentiel dans le prix de la capacité est une quasi-incitation à essayer d'exercer un pouvoir de marché, puisqu'une petite variation des quantités sur le marché peut être extrêmement rémunératrice (ce phénomène n'est en rien propre aux marchés électriques, mais il est fortement aggravé par leurs spécificités).

Le deuxième problème est le prix plafond de la capacité imposé de facto par la pénalité. En effet, Joskow (2006b) remarque que, sur plusieurs années, le niveau de capacité objectif sera atteint certaines années et ne le sera pas pour d'autres. Or, le fait d'avoir fixé la pénalité au niveau qui permet de rémunérer les coûts fixes d'une unité de pointe (sur une journée) implique que, en moyenne, les revenus dégagés du marché de capacité ne permettront pas de couvrir tous les coûts fixes sur plusieurs années (ce qui explique son constat que même sur le PJM, il semble exister un problème de « revenus manquants »). Cependant, ces revenus n'étaient censés être qu'un complément aux rentes de rareté dégagées sur le marché spot. Même si elle reste purement théorique au vu de l'ampleur du problème des « revenus manquants », la possibilité existe que les producteurs reçoivent plus que leurs coûts fixes quand ils combinent rentes de rareté et paiements de capacité.

Ces problèmes expliquent que de nouveaux designs (ainsi qu'on appelle l'ensemble des règles et institutions permettant le fonctionnement d'un marché) de marché de capacité aient vu le jour ces dernières années, certains restant théoriques, d'autres sont déjà en voie d'implémentation. Parmi ces nouveaux designs, deux méritent particulièrement d'être évoqués : le LICAP (pour Locational Installed Capacity) proposé en 2004 en Nouvelle-Angleterre (et tout à fait similaire aux recommandations de la CPUC dans son « Capacity Markets White Paper » de 2005 ou aux analyses de Cramton et Stoft dans leur article de 2005) et le FCM (Forward Capacity Market) proposé par Cramton et Stoft (2006) et repris par Joskow (2006b). Le FCM est en cours de négociation en Nouvelle-Angleterre. Ces deux designs sont considérés par Cramton et Stoft comme faisant partie des designs « convergents ». Ils distinguent, en effet, parmi les designs proposés pour remédier aux problèmes de sécurité d'approvisionnement, entre ceux qui reposent sur les marchés de capacité « traditionnels » (c'est-à-dire essentiellement les marchés dit ICAP existant dans le Nord-est des Etats-Unis) et ceux qui reposent essentiellement sur des réformes des marchés d'énergie (« Energy-Only »). Ces deux familles de designs proviennent des débats entre économistes de l'électricité et peut recouper une distinction entre les avocats de

solutions « par le marché », pour lesquels les marchés de capacités sont une structure administrative qui ne peut qu'aggraver les problèmes de sécurité d'approvisionnement (puisque les marchés de capacités sont centralisés et comportent tous, plus ou moins, une gestion administrative de la demande ; Wolak (2004) et Hogan (2005) en sont deux exemples) ; et ceux pour lesquels, en l'état actuel, les marchés d'électricité connaissent de trop fortes défaillances pour qu'il ne faille pas bâtir des structures ad hoc pour y remédier (Stoft en est un avocat à travers l'analyse développée dans son ouvrage « Market Power Economics » de 2002).

Cramton et Stoft (2006) se font donc forts de combiner le meilleur des deux approches dans ce qu'ils appellent les designs convergents (et dont ils sont pour une grande part à l'origine, tant théoriquement que pratiquement via leurs activités de consultants). On peut sommairement résumer leur argumentation en expliquant que l'avantage des marchés de capacités « traditionnels » est de fixer des objectifs de capacités explicites qui conditionnent l'ampleur des revenus additionnels censés remédier au problème des revenus manquants (qu'ils sont donc les seuls à traiter explicitement). Tandis que l'avantage des approches « energy only » est de proposer des moyens de mieux couvrir les risques d'investissement par le biais de mécanismes optionnels (mais sans reconnaître, ni essayer de traiter, l'importance des « revenus manquants »).

Les analyses proposées dans les précédentes parties de cette thèse ont clairement montré que les mécanismes de prix sur le marché spot ne pouvaient constituer des remèdes efficaces au problème des revenus manquants (problème non seulement reconnu par Joskow (2006a et 2006b), Joskow et Tirole (2006), mais également par Hogan (2005)). En effet, l'existence d'un pouvoir de marché par l'investissement, ainsi que le statut particulier des réserves opérationnelles (dont les coûts ne peuvent être intégrés à travers les prix du marché spot), sont des arguments centraux pour essayer de construire des politiques de restauration des incitations à l'investissement à travers la fixation explicite d'objectifs de capacités, plutôt que par des mécanismes impliquant le prix spot. C'est justement ce

qu'introduisent explicitement les designs du LICAP et du FCM. La capacité est alors rémunérée de façon séparée de l'énergie par le biais d'un marché ad hoc.

Un autre argument expliquant l'intérêt de cette fixation est avancé dans le « Capacity Market White Paper » de la CPUC (2005) : le meilleur remède au pouvoir de marché (au sens traditionnel) est de s'assurer de l'absence de déficit de capacité. En effet, tout déficit induit que un ou quelques producteurs sont à même de fixer les prix à la pointe, puisqu'ils sont indispensables pour éviter l'effondrement du système. Fixer un objectif de capacité permet de fortement diminuer le lien entre prix de l'énergie et niveau des capacités de production.

Le LICAP

Le LICAP peut être vu comme une version plus rudimentaire du FCM. Il est donc logique d'en présenter les mécanismes en premier, dans la mesure où leur compréhension permet d'appréhender les mécanismes du FCM (qui sont plus complexes). La première innovation qu'apporte le LICAP, ainsi qu'on l'a déjà cité, consiste en l'utilisation d'une courbe de demande de capacité administrative pentue et non pas verticale. Il s'agit d'éviter l'exercice d'un pouvoir de marché sur le marché de capacité, tout en rendant possible une plus grande stabilité des revenus issus de ce marché. Toute la demande de capacité est résumée dans cette courbe : il n'y a pas de confrontation entre offre et demande (ce qui supprime la possibilité d'un pouvoir de marché au sens traditionnel). La courbe est paramétrée de telle sorte que le prix correspondant au niveau de capacité objectif est égal aux coûts fixes d'une unité de pointe. Tous les types d'unités de production (y compris les unités de base) sont éligibles pour recevoir les paiements. Le raisonnement sous-jacent est en tout point similaire à celui de Stoft (2002) avec les fonctions de profit : si le prix est trop élevé, il y aura sur-investissement, ce qui fera baisser le prix et inversement. En fait, la fonction de profit est imposée administrativement. Le prix issu de la confrontation de la courbe de demande administrative avec la

courbe d'offre agrégée devient la pénalité pour non-respect des objectifs de capacité, évitant ainsi de plafonner le prix de la capacité.

Ce fonctionnement rappelle celui du défunt Pool anglais qui centralisait toutes les offres d'énergie et les confrontait avec une prévision de la demande. De même, le Pool incluait une rémunération de la capacité dans le prix de l'énergie, rémunération calculée en effectuant le produit du coût de la défaillance par la probabilité de celle-ci (appelée LOLP, pour Loss Of Load Probability). Ce mécanisme a donné lieu à de nombreuses manipulations, notamment sur la rémunération de la capacité, puisque la LOLP était extrêmement sensible à des retraits de capacité en situation de tension offre-demande.

Le LICAP évite cet écueil par le biais de deux dispositions : la première, est que c'est la capacité installée totale qui fixe le niveau du prix, et non pas les offres sur le marché (qui ne déterminent que l'identité des offreurs sur une base mensuelle). Cela supprime toute possibilité de faire monter le prix par le retrait de capacités.

La deuxième réside dans la soustraction des rentes de rareté (appelées Peak Energy Rents dans les documents présentant ce design) provenant des pics de prix sur le marché de l'énergie aux revenus dégagés par le marché de capacité. Ainsi est évité le risque que la combinaison des deux flux n'induisse des revenus supérieurs aux coûts fixes (ce qui pourrait conduire à un sur-investissement et donc à des cycles d'investissement). Mais cela rend également possible la réduction du risque des investisseurs, dans la mesure où les paiements de capacités agissent comme une assurance contre de trop faibles rentes de rareté les années de faible demande (et leur permettent de planifier les flux de revenus d'un investissement). La rémunération des investissements est lissée : les années de fortes demandes verront des revenus dégagés par les pics de prix ; tandis que les années de faibles demandes verront les revenus essentiellement issus du marché de capacité. Ce mécanisme est parfaitement cohérent avec la volonté de combler les « revenus manquants » et il supprime l'intérêt à exercer un pouvoir de marché au sens traditionnel sur le marché de l'énergie, puisque toute hausse des rentes de rareté se traduit par des paiements de capacités moindres. De plus, les PER sont calculées

ex-post, les paiements de capacités peuvent donc être ajustés de façon précise pour atteindre le niveau des coûts fixes. Ces deux dispositions sont les deux innovations importantes du LICAP par rapport au design « traditionnel ».

Un dernier changement complète le dispositif : les paiements de capacités ne sont attribués que proportionnellement à la capacité des offreurs à produire pendant les heures où existe un déficit de capacité (par exemple, de réserves opérationnelles ou bien les heures pendant lesquelles le prix spot est supérieur au coût marginal du système). Ainsi, un mécanisme d'incitation à la performance est introduit dans le design.

Revenons sur la deuxième disposition. Cette façon de calculer les paiements de capacité induit qu'un producteur dont toute la capacité est offerte sur le marché de capacité ne dépend même plus des rentes de rareté. Son revenu total est en effet égal aux rentes infra-marginales (nulle pour une unité de pointe) auxquelles s'ajoutent les paiements de capacité (avant soustraction des PER). Ces derniers remplacent complètement les rentes de rareté. Mieux, c'est bien la somme de ces deux quantités qui est effectivement payée par les consommateurs. On est bien en présence d'une gestion administrative de la pointe (puisque ce sont les rentes de rareté qui sont remplacées, pas les rentes infra-marginales), sans pour autant supprimer les mécanismes de marché, la seule existence du problème des « revenus manquants » étant une motivation suffisante pour que les producteurs souhaitent participer au marché de capacité. Les incitations à investir semblent être restaurées par ce simple mécanisme. Dès lors, un pouvoir de marché par l'investissement ne peut plus s'exercer qu'à travers la possibilité éventuelle d'influer sur les paiements de capacité. Autrement dit, cette innovation est particulièrement intéressante à la lumière des analyses de la deuxième partie. A la pointe, partie de la demande la plus sensible à l'exercice d'un tel pouvoir, les seuls revenus qui permettent de couvrir les coûts fixes sont les paiements de capacité. Si ces derniers ne sont pas sensibles à ce pouvoir, alors il est possible de résoudre ce problème, tout en empêchant également l'exercice d'un pouvoir de marché traditionnel.

Cependant, la courbe de demande administrative du LICAP, si elle supprime tout pouvoir de marché traditionnel, reste soumise à l'exercice d'un pouvoir de marché par l'investissement. En effet, si l'on suppose que la structure du marché est oligopolistique, cela induit que les firmes vont choisir leur niveau de capacité de telle sorte qu'il maximise les paiements de capacités sur la courbe de demande. Il y a un fort risque de sous-investissement par rapport à l'objectif et de sur-rémunération de la capacité.

Le FCM

Le FCM reprend en les affinant les principales innovations du LICAP et introduit deux innovations : des marchés de capacités forward et un mécanisme optionnel, permettant à la fois de couvrir le risque des investisseurs et de créer de fortes incitations à la performance par le biais de prix « magnifiés » (c'est-à-dire multipliés par une constante au-delà d'un certain niveau) dans la rémunération de cette dernière (ou plutôt de la « sur-performance par rapport à une unité de référence).

En revanche, il supprime la courbe de demande administrative, dont on a vu qu'elle risquait d'induire des comportements stratégiques. Le prix de la capacité (et donc les paiements) est fixé par une enchère dont la demande est inélastique et égale à l'objectif de capacité à atteindre. Dans la mesure où les producteurs déjà existants peuvent enchérir à des prix inférieurs à leurs coûts fixes, ce sont les enchères des nouveaux entrants qui vont fixer le prix de la capacité. D'où l'intérêt que le marché soit un marché forward. En effet, les marchés de capacité forward (qui vont jusqu'à 4 années dans le futur dans les propositions de Cramton et Stoft (2006)) permettent de coordonner l'entrée de nouvelles capacités, par la fixation d'objectifs futurs de capacité. Ce dispositif peut être un moyen d'éviter les cycles d'investissement tels qu'ils ont été souvent observés dans différents systèmes électriques déréglementés. La possibilité de s'assurer du niveau futur des revenus

générés par un investissement (sur les premières années du projet) représente également un moyen de faciliter l'arrivée de nouveaux entrants et donc de limiter le pouvoir de marché par l'investissement.

Enfin, Cramton et Stoft (2006) recommandent que les enchères qui fixent le prix de la capacité soit des enchères descendantes, dites « descending clock auction ». Dans le cadre de ces dernières, les offreurs n'enchérissent que des quantités. De plus, l'enchère débute à un prix élevé et tous les offreurs doivent enchérir toute leur capacité à cet instant. Puis, le prix baisse à intervalles réguliers. A chaque révision du prix les offreurs peuvent maintenir leur offre, la diminuer ou bien se retirer de l'enchère. Cramton et Stoft précisent également les différentes dispositions susceptibles de limiter l'exercice d'un pouvoir de marché dans la fixation du prix de la capacité, puisque la seule possibilité encore ouverte à un tel pouvoir dans le design se trouve dans cette enchère. Ils recommandent en particulier de ne pas accepter de retraits des producteurs déjà en place à l'exception des mises sous cocon, des démantèlements ou des exports. Ce faisant, les risques qu'ils puissent faire monter le prix auquel sont rémunérées leurs capacités non retirées sont fortement diminués.

L'ensemble des dispositifs qui composent ce design est relativement bien pensé pour éviter des comportements stratégiques et notamment le pouvoir de marché par l'investissement. A chaque fois que c'est possible, les fixations conjointes des prix et des quantités sont évitées afin de parvenir au meilleur réglage des résultats attendus du design. Celui-ci représente donc une réelle avancée par rapport à l'ensemble des dispositifs pensés pour restaurer les incitations à investir.

Cependant, la possibilité pour un investisseur de pouvoir complètement couvrir ses risques pose plusieurs problèmes. Tout d'abord, il convient de se demander qui supporte le risque ainsi transféré. L'analogie avec les options constamment évoquée par Cramton et Stoft renforce cette interrogation. Sur un marché d'options, les acheteurs d'options payent une prime (le prix d'achat) pour se

couvrir contre un risque donné. Mais cela n'est possible que parce qu'il y a des vendeurs en face qui sont prêts à supporter ce risque moyennant la prime. Or, le transfert de risque dans ce design est peu clair, car les consommateurs ne sont pas censés payer plus que le prix de capacité (qui, lorsque l'objectif de capacité est atteint, doit être égal au coût fixe d'une unité de pointe de référence).

Un autre point important mérite d'être soulevé : comment sont rémunérées les réserves opérationnelles dans ce schéma ? L'article de Cramton et Stoft (2006) ne présente pas de façon complète le design du FCM, néanmoins aucune mention des réserves opérationnelles n'est faite. Or, comme nous l'avons montré, le problème de sécurité des approvisionnements ne se résume pas à faire en sorte que suffisamment de capacités de production soient construites. Il est également nécessaire qu'existent suffisamment de moyens utilisables comme réserves opérationnelles. Ces dernières doivent en effet être suffisamment réactives pour permettre au système électrique de faire face à des variations subites de demande. De même, la localisation des différents moyens utilisables comme réserves est un paramètre important pour pouvoir les utiliser à bon escient pour soutenir le réseau de distribution. Or, dans le design du FCM toutes les capacités sont apparemment rémunérées indépendamment de la prise en compte de ces aspects (ou d'autres). D'où un problème potentiel : qu'il existe des capacités installées suffisantes, mais trop peu de moyens éligibles comme réserves.

Or, dans le modèle de la partie 4 de l'article de Joskow et Tirole (2006), l'égalité qui décrit le programme d'investissement optimal fait apparaître d'un côté les coûts fixes d'un générateur, et de l'autre, deux quasi-rentes. L'une, liée à l'accroissement de surplus social que procure la diminution de la probabilité de délestage apportée par une unité supplémentaire de capacité. L'autre, correspondant à l'accroissement du surplus social que procure la diminution de la probabilité d'un effondrement du réseau apportée par une unité supplémentaire de capacité (dans la zone des états du modèle où le taux de réserve est variable). Or, si

ces quasi-rentes dépendent bien de la valeur des capacités installées à travers les probabilités de se retrouver dans la zone de délestage ou dans la zone de réduction des réserves, elles dépendent également des états de la nature du modèle, autrement dit des réalisations de la demande. L'égalité entre coûts fixes et quasi-rentes n'est valable qu'en espérance. Alors que le mécanisme de couverture semble abolir tout lien (pour les capacités complètement « couvertes » par leur participation aux marchés de capacité forward) entre rémunération des investissements et les aléas qui gouvernent les systèmes électriques.

Il est possible d'illustrer ce point d'une manière plus explicite en reprenant un certain nombre de spécifications du modèle de Joskow et Tirole (2006), notamment les mécanismes qui déterminent la possibilité d'un effondrement du système ; mais en introduisant deux types de générateurs, dont un est éligible comme réserve opérationnelle et l'autre non (du fait des vitesses de réaction de chacun par exemple).

ii. Un modèle de réserves opérationnelles

Il est extrêmement difficile de décrire le programme d'investissement d'un producteur électrique dans le cadre élaboré par Joskow et Tirole. Nous allons plutôt nous intéresser aux résultats du programme d'investissement optimal d'un planificateur, afin de distinguer plusieurs types de revenus fictifs susceptibles de servir comme points de référence.

Les deux types de générateurs ont des coûts d'investissements de I_1 et I_2 , avec I_2 strictement supérieur à I_1 ; le générateur éligible comme réserve opérationnelle étant celui qui a les coûts fixes les plus importants (cela fin d'éviter la solution triviale où tout le parc de production est mono-technologique). La demande est inélastique et aléatoire : dans l'état de la nature i (indiqué sur $[0,1]$), elle vaut : D_i . Si une quantité $d_i \leq D_i$ est servie, le surplus des consommateurs sera égal à $d_i C_D$, où C_D est le coût de la défaillance. On suppose qu'il existe un deuxième aléa λ affectant la disponibilité des capacités totales (qui prend donc ses valeurs sur $[0,1]$), indépendant de la demande. Sa densité est notée h et sa distribution H . On fait l'hypothèse que :

$\frac{h(\lambda)\lambda}{1-H(\lambda)}$ est une fonction croissante de λ .

La séquence des décisions est la même que dans le modèle de Joskow et Tirole (2006) : tout d'abord, le niveau de capacité pour chaque type de générateur est fixé, puis la réalisation D_i de la demande est connue. Le dispatcher choisit alors de servir une quantité d_i de la demande et détermine le taux de réserve r_i , de telle sorte que : $r_i d_i \leq K_2$. Le coût lié à l'utilisation d'une unité de réserves est s . Enfin, la réalisation de λ dans l'état de la nature i est connue et, soit la part de la demande d_i est servie, soit le système s'effondre. Cette dernière éventualité est donc réalisée quand on a : $\lambda_i(1+r_i)d_i \leq d_i$, soit : $\lambda_i(1+r_i) \leq 1$

L'interprétation du modèle est la suivante : le deuxième aléa (celui qui peut faire s'effondrer le réseau) nécessite l'intervention de générateurs très rapides (par exemple démarrant en moins de 10 minutes) donc de générateurs du deuxième type, là où les générateurs du premier type ont un temps de démarrage plus long. Cela ne signifie pas qu'ils ne peuvent pas être utilisés comme réserves opérationnelles, mais plutôt qu'il existe plusieurs types d'aléas post dispatch qui nécessitent des moyens plus ou moins rapides. Le modèle ne comporte néanmoins qu'un seul aléa afin d'en simplifier la résolution. Il n'est donc pas possible de garder le niveau de généralité du modèle de Joskow et Tirole (2006), les réserves opérationnelles décrites ici ne servant qu'à répondre à l'aléa post dispatch. Cela modifie les contraintes dans la maximisation du planificateur, puisqu'on ne suppose plus que les capacités doivent être supérieures à la demande multipliée par un plus le taux de réserve, mais seulement supérieures à la demande.

Le surplus social est égal à l'espérance de satisfaction des consommateurs moins les coûts d'investissement dans les différents types de générateurs. La demande satisfaite est valorisée au coût de la défaillance moins le coût variable de production.

La maximisation du surplus social est donc équivalente à :

$$\max_{K_1, K_2, d_i, r_i} \left\{ E \left[\left[1 - H \left(\frac{1}{1 + r_i} \right) \right] (C_D - C_P) - s r_i \right] d_i - I_1 K_1 - I_2 K_2 \right\}$$

sous des contraintes telles que pour tout état i :

$$d_i \leq D_i \quad (\alpha_i)$$

$$r_i d_i \leq K_2 \quad (\beta_i)$$

$$d_i \leq K_1 + K_2 \quad (\gamma_i)$$

Les variables entre parenthèses étant les multiplicateurs associés aux contraintes.

Dans la mesure où le planificateur préférera baisser le taux de réserves plutôt que délester, la première contrainte peut être supprimée car il n'y aura de délestages que si la demande est supérieure au total des capacités. Les contraintes deviennent alors :

$$r_i d_i \leq K_2 \quad (\beta_i)$$

$$d_i \leq K_1 + K_2 \quad (\gamma_i)$$

Les conditions de premier ordre sont alors (pour respectivement r_i , d_i , K_1 et K_2) :

$\frac{h}{(1+r_i)^2} (C_D - C_P) - s = \beta_i$	(1)
$\left[1 - H\left(\frac{1}{1+r_i}\right) \right] (C_D - C_P) - s r_i = r_i \beta_i + \gamma_i$	(2)
$E(\gamma_i) = I_1$	(3)
$E(\beta_i + \gamma_i) = I_2$	(4)

Les deux premières relations vont définir les régimes du modèle; les deux dernières, la politique optimale d'investissement.

Notons que γ_i ne peut être nul, quand β_i ne l'est pas, puisque alors l'espérance de γ_i est égale à deux valeurs strictement différentes, ce qui est impossible. Donc, la proposition suivante est vraie :

$$(\gamma_i = 0) \Rightarrow (\beta_i = 0)$$

Quand la dernière contrainte ne joue pas, la deuxième non plus. Ou si l'on reformule en termes économiques cette implication : il ne peut pas y avoir des réserves non contraintes par les capacités de production, lorsque ces dernières sont inférieures au niveau de demande issu du dispatching. La présence des deux types de générateurs ne change rien à la stratégie optimale qui consiste à d'abord faire baisser le niveau des réserves plutôt que délester une partie de la demande.

Régimes de fonctionnement

Quand la demande est basse relativement aux capacités, toute la demande est servie ($d_i = D_i$), et les multiplicateurs γ_i et β_i sont nuls. On peut parler d'une zone de surcapacité. De (1), on tire donc :

$\frac{h\left(\frac{1}{1+r_i}\right)}{(1+r_i)^2} C_D - s = 0$, le taux de réserve est alors constant et égal à r_H , puisque les hypothèses posées garantissent l'unicité de la solution de cette équation.

Cette zone est définie par la conjonction des deux inégalités :

$$D_i \leq K_1 + K_2$$

$$r_H D_i \leq K_2$$

Les revenus fictifs des générateurs dans cette zone sont nuls. Dans la mesure où les deux dernières contraintes ne jouent pas, il n'y a aucune augmentation du surplus du fait d'une augmentation de la capacité d'une des deux technologies. Nous avons

déjà montré que quand la demande augmente de telle sorte qu'on change de zone, ce sont les réserves qui vont d'abord être baissées de préférence à un délestage.

La deuxième zone de fonctionnement du modèle est celle pour laquelle la demande est supérieure aux capacités et dans laquelle des délestages sont opérés. Elle voit le taux de réserves décroître au fur et à mesure que la demande augmente.

Le programme d'investissement optimal

Le programme d'investissement optimal est donné par les deux dernières conditions du premier ordre. Le premier type de générateur ne perçoit qu'une quasi-rente, celle correspondant à la diminution de la probabilité de défaillance qu'il permet (associée au quasi-prix γ_i). Le deuxième type perçoit ce revenu auquel il faut ajouter la quasi-rente associée à sa contribution à la baisse de la probabilité d'effondrement du système (associée au quasi-prix β_i).

La première quasi-rente s'écrit :

$$I_1 = \int_{K_1+K_2}^{+\infty} \left[\left(1 - H\left(\frac{1}{1+r_i} \right) \right) (C_D - C_P) - \frac{hr_i}{(1+r_i)^2} (C_D - C_P) \right] f_i di$$

Et la deuxième :

$$I_2 = \int_{K_1+K_2}^{+\infty} \left[\left(1 - H\left(\frac{1}{1+r_i} \right) \right) (C_D - C_P) - (1-r_i) \frac{h}{(1+r_i)^2} (C_D - C_P) \right] f_i di$$

D'une certaine façon, la première quasi-rente correspond au prix de la capacité, et le quasi-revenu total perçu par le deuxième-type de générateur correspond quant à lui au prix des réserves opérationnelles. Le deuxième terme de son revenu est la partie qui rémunère la vitesse de réaction seule à même faire face au deuxième aléa du modèle. Ces revenus ne prennent des formes aussi nettes, qu'en raison de l'indépendance des deux aléas du modèle. Or, dans la réalité, la loi de probabilité d'un effondrement du réseau est corrélée avec celle de la demande. C'est une des raisons pour lesquelles il est extrêmement difficile de valoriser les réserves opérationnelles, et plus généralement les contraintes dynamiques qui affectent la plupart des générateurs. Néanmoins, ce modèle nous permet d'obtenir une première approche de la rémunération de la vitesse de réaction des générateurs (à travers leur contribution à réduire la probabilité de l'effondrement du réseau).

Là où ces résultats présentent un intérêt particulier, c'est ce qu'ils nous permettent d'entrevoir dans le mécanisme de couverture du FCM. En effet, le prix de capacité du FCM ne rémunère que ... la capacité. Il n'inclut pas de rémunération de la capacité des générateurs à remplir des missions de réserves opérationnelles. Or, dans le cas exposé dans le modèle, la non-prise en compte de cet aspect conduirait à n'investir que dans le premier type de générateur, au détriment de la stabilité du réseau. Bien qu'hypothétique, il reflète les problèmes que pose la rémunération de la capacité indépendamment des autres caractéristiques nécessaires au bon fonctionnement d'un système électrique. Dans la mesure où ces caractéristiques peuvent être plus ou moins assurées par plusieurs types de générateurs différents, le risque existe avec le FCM de se retrouver avec un parc de production possédant le bon niveau de puissance sans pour autant disposer des différents moyens nécessaires comme réserves opérationnelles.

Il existe des possibilités pour un gestionnaire de réseau de rémunérer ces moyens en dehors des voies habituelles par le biais de dispatch ne respectant pas le « merit order ». Mais, le problème des incitations à l'investissement pour ces moyens demeure, puisqu'il est difficile pour un investisseur de prévoir les besoins futurs du

gestionnaire. Par ailleurs, un tel dispatch pose de sérieux problèmes de distorsion des prix et peut donc éventuellement conduire à déprimer les incitations à investir. De plus, il est difficile d'organiser des marchés différents suivant les différentes caractéristiques nécessaires. D'abord, parce qu'on sait mal les valoriser (à l'instar des contraintes dynamiques) ; ensuite, parce qu'il est difficile d'organiser de tels marchés autour de volumes réduits : les besoins en capacité rapides, par exemple, dépendent de la topologie du réseau considéré, ce qui nécessite de tenir compte de la localisation des différents moyens par le biais d'autant de sous marchés potentiels qui vont manquer de la profondeur nécessaire à un bon fonctionnement. De même, les besoins en réserves opérationnelles peuvent ne représenter qu'un faible volume de transaction.

Enfin, un dernier problème s'oppose à une valorisation des réserves par le biais d'un marché ad hoc, c'est la forte instabilité du prix de l'énergie dès lors qu'apparaît un déficit des réserves. Dans cette situation, les fonctions d'offre et de demande d'énergie sont verticales et confondues (Joskow, Tirole 2006). Or, il existe une relation d'arbitrage plus ou moins forte entre le prix de l'énergie et le prix des réserves. Ainsi, la moindre fluctuation sur l'un des marchés va se traduire par de forts mouvements des prix, étant entendu que ces fluctuations dépendent en grande partie du comportement du gestionnaire de réseau et de sa politique de gestion d'un déficit de réserves. Ce problème, dit du « knife edge », va à l'encontre d'une gestion efficace des réserves de la part du gestionnaire de réseau, tout en ajoutant une difficulté supplémentaire importante (voire définitive) à ce que des investisseurs disposent des incitations suffisantes à investir. Tous ces problèmes plaident pour un traitement plus radical des réserves opérationnelles.

c-La propriété publique des moyens de pointe

Les problèmes que posent le statut de bien public des réserves opérationnelles et la difficulté d'intégrer les coûts sociaux liés à la pointe, peuvent plaider en faveur de la propriété publique des moyens de production de pointe. En effet, si ces moyens sont construits et opérés par le gestionnaire de réseau, le choix adéquat des technologies et des localisations garantit un niveau de réserves opérationnelles optimal. Le statut de bien public des réserves et les difficultés déjà soulignées à financer celles-ci semblent imposer une telle solution : il n'est pas possible a priori de différencier les paiements de capacité suivant la possibilité qu'auraient les moyens financés à jouer un rôle comme réserves, et cela du fait de la multiplicité des critères à prendre en compte. La propriété publique permet également de tenir compte du coût social d'une insuffisance de capacités engendrant des délestages ou un effondrement, puisque les moyens de pointe sont en général les moyens éligibles comme réserve. Il semble donc possible de remédier aux différents problèmes posés précédemment. Les segments inférieurs dans le « merit order » sont par contre toujours soumis à une concurrence entre plusieurs producteurs.

Le problème des « revenus manquants » existe donc toujours, puisque les rentes de rareté constituent une partie des revenus dégagés par tous les moyens de production, et ce, même si leur importance décroît au fur et à mesure qu'on s'approche de la base. Il est donc nécessaire de coupler la propriété publique des moyens de pointe à un dispositif assurant une rémunération de la capacité, comme le FCM. Ce dernier reste le plus attrayant des dispositifs proposés puisqu'il est le plus susceptible de réduire les possibilités d'exercice d'un pouvoir de marché (par l'investissement ou la restriction de la production). En effet, en faisant des paiements de capacité un complément aux rentes de rareté dégagées sur le marché de l'énergie, toute manipulation des prix de l'énergie visant à augmenter les profits, se traduit par une baisse équivalente des paiements de capacité.

Si Meunier et Finon (2006) se préoccupent de la possibilité pour le gestionnaire de réseau de posséder et opérer en propre des moyens de production, ils ne l'envisagent que comme une menace en vue d'inciter à investir suffisamment les producteurs disposant d'un pouvoir de marché par l'investissement. Or, si cette menace permet d'obtenir le niveau d'investissement optimal, les producteurs sont toutefois susceptibles de retenir un surprofit en restreignant l'utilisation de leurs capacités de production lors de cette étape. Le FCM apparaît donc comme plus adéquat pour contrôler le pouvoir de marché des producteurs.

Une esquisse de dispositif se profile donc, combinant FCM et propriété publique des moyens de pointe, afin d'assurer un fonctionnement optimal des marchés électriques. Le modèle qui suit a pour but d'essayer de valider cette intuition. Il reprend le cadre développé dans la deuxième partie, en y incorporant une différenciation entre deux types de moyen de production : une base soumise à concurrence et une pointe construite et opérée par le gestionnaire de réseau.

Spécifications du modèle

Nous reprenons le cadre d'analyse développé dans la partie 2, mais en supposant qu'existent deux types de moyens de production : l'un assurant la pointe et apte à être utilisé comme réserves ; l'autre assurant la fourniture de base. Le premier est construit et opéré par le gestionnaire de réseau (ISO), tandis que le deuxième est construit et opéré par un oligopole en concurrence. Nous supposons que le prix est égal au coût marginal du système, sauf quand la demande est supérieure à la capacité installée, auquel cas le prix est égal au prix plafond. Autrement dit, tant que la demande est inférieure aux capacités de production, une concurrence à la Bertrand oppose les producteurs ; quand la demande est supérieure, les producteurs produisent à pleine capacité.

Nous supposons également que le gestionnaire de réseau va investir jusqu'à ce que les capacités installées soient égales au niveau optimal qui permet d'égaliser le coût de construction d'une unité supplémentaire de capacité de pointe avec le coût espéré d'un mégawattheure de défaillance, ainsi qu'il est exposé dans la première partie.

Le déroulement du jeu entre les producteurs et le gestionnaire de réseau est donc le suivant : dans un premier temps, la demande n'est pas connue et les producteurs choisissent leurs capacités de production de base ; dans un deuxième temps, le gestionnaire de réseau investit en moyens de pointe afin d'atteindre le niveau de capacité totale optimale ; puis la demande est connue, les firmes fournissent la demande au prix résultant de la confrontation entre demande et capacités.

La comparaison entre l'optimum de Ramsay et les solutions du jeu entre les producteurs et le gestionnaire de réseau permet de savoir s'il est possible d'atteindre l'optimum. Il faut donc calculer dans un premier temps l'optimum de Ramsay, puis résoudre le programme d'investissement d'un producteur.

L'optimum de Ramsey

Le niveau optimal des capacités est donné par l'équation :

$$P(D \geq \bar{Q}_{tot}^*) = \frac{I_2}{(C_D - C_{P_2})}$$

On retrouve ici le résultat présenté dans la première partie.

La répartition entre la capacité de production de base et la capacité de pointe est donnée par l'égalité :

$$P(D \geq \bar{Q}_1) = \frac{I_1 - I_2}{C_{P_2} - C_{P_1}}$$

La capacité de base est construite jusqu'au point où il devient moins cher (en espérance) de produire avec la technologie de pointe. La capacité de pointe se déduit en faisant la différence entre le niveau de capacité total donné par la première équation et la capacité de base donnée par la deuxième. Cette répartition permet de minimiser le coût total de production.

Le programme d'investissement d'un producteur

Le choix d'investissement d'une firme est déterminé par la maximisation de son espérance de profit, qui s'écrit :

$$E[\pi_i] = E \left[-I_1 \bar{Q}_i + (C_{P_2} - C_{P_1}) \bar{Q}_i 1_{\left\{ \sum_{j=1}^n \bar{Q}_j < D < \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right\}} + (\bar{P} - C_{P_1}) \bar{Q}_i 1_{\left\{ D \geq \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right\}} \right]$$

Soit aussi :

$$E[\pi_i] = -I_1 \bar{Q}_i + (C_{P_2} - C_{P_1}) \bar{Q}_i P \left(\sum_{j=1}^n \bar{Q}_j < D < \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right) + (\bar{P} - C_{P_1}) \bar{Q}_i P \left(D \geq \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right)$$

$$\begin{aligned} E[\pi_i] = & -I_1 \bar{Q}_i + (C_{P_2} - C_{P_1}) \bar{Q}_i P \left(D < \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right) - (C_{P_2} - C_{P_1}) \bar{Q}_i P \left(D < \sum_{j=1}^n \bar{Q}_j \right) \\ & + (\bar{P} - C_{P_1}) \bar{Q}_i P \left(D \geq \sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 \right) \end{aligned}$$

Cependant, la capacité totale est une constante, car les producteurs savent que le gestionnaire de réseau va construire jusqu'à ce que soit atteint le niveau de capacité totale optimal. On a donc :

$$\sum_{j=1}^n \bar{Q}_j + \bar{Q}_2 = \bar{Q}_{tot}^*$$

En tenant compte de cette égalité, la condition de premier ordre nous donne la fonction de réaction de la firme i :

$$\begin{aligned} \frac{\partial E[\pi_i]}{\partial \bar{Q}_i} = & -I_1 + (C_{P_2} - C_{P_1})P(D < \bar{Q}_{tot}^*) - (C_{P_2} - C_{P_1})P\left(D < \sum_{j=1}^n \bar{Q}_j\right) \\ & + (\bar{P} - C_{P_1})P(D \geq \bar{Q}_{tot}^*) - (C_{P_2} - C_{P_1})\bar{Q}_i \frac{\partial P\left(D < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i} = 0 \end{aligned}$$

L'équilibre est symétrique, comme permet de le montrer la différence des fonctions de réaction les unes avec les autres.

La condition de deuxième ordre est alors :

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = -2(C_{P_2} - C_{P_1}) \frac{\partial P\left(D < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i} - (C_{P_2} - C_{P_1})\bar{Q}_i \frac{\partial^2 P\left(D < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i^2}$$

On se place à l'équilibre, on a donc : $\sum_{j=1}^n \bar{Q}_j = n\bar{Q}$

Or :

$$\frac{\partial P\left(D < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i} = \frac{\partial P\left(M + \varepsilon < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\sum_{j=1}^n \bar{Q}_j - M\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(n\bar{Q} - M)^2}{2\sigma^2}}$$

$$\frac{\partial^2 P\left(D < \sum_{j=1}^n \bar{Q}_j\right)}{\partial \bar{Q}_i^2} = -\frac{1}{2\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(n\bar{Q} - M)^2}{2\sigma^2}} (2n^2\bar{Q} - 2nM)$$

Donc, la condition de deuxième ordre peut se réécrire :

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = -(C_{P_2} - C_{P_1}) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(n\bar{Q} - M)^2}{2\sigma^2}} \left(2 - \frac{1}{2\sigma^2} \bar{Q} (2n^2\bar{Q} - 2nM)\right)$$

$$\frac{\partial^2 E[\pi_i]}{\partial \bar{Q}_i^2} = -(C_{P_2} - C_{P_1}) \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(n\bar{Q} - M)^2}{2\sigma^2}} (-n^2\bar{Q}^2 + nM\bar{Q} + 2\sigma^2)$$

Le signe de cette dérivée est donné par l'opposé de celui du polynôme (puisque tous les autres termes sont positifs). Nous retrouvons l'intervalle de définition des résultats de la partie 2.

Si l'on compare le niveau d'investissement de l'oligopole à celui optimal, nous constatons qu'il n'y a aucune raison pour que les deux soient égaux. En effet, là où l'optimum tient compte des coûts des différentes technologies pour déterminer la répartition entre les deux, la maximisation du profit d'une entreprise de l'oligopole n'intègre pas le coût d'investissement dans la technologie de pointe. De plus, les producteurs sont toujours à même d'influer sur leur distribution de profit par le biais du choix de leur capacité de production : ce n'est plus la rente de rareté qui est modifiée, mais la rente infra-marginale. Autrement dit, même si les risques de manipulation des rentes de rareté sont supprimés par le biais d'un dispositif comme le FCM, subsistent des possibilités pour les producteurs (même si elles sont

moindres) d'exercer un pouvoir de marché. Dès lors, l'optimum de second rang n'est pas atteignable.

Ce résultat négatif souligne également les difficultés à traiter séparément les différents segments de production (pointe, base, etc.). Toute intervention à un niveau va nécessairement avoir des conséquences sur les autres. Dans un cadre assez proche, Joskow et Tirole (2006 partie 3) indiquent d'ailleurs qu'il n'est pas possible d'atteindre l'optimum de second rang à l'aide d'un prix plafond (empêchant l'exercice d'un pouvoir de marché) et d'un paiement de capacité s'il existe plusieurs segments de production et plus de trois états de la nature. Le prix plafond permet de corriger le pouvoir de marché pour les moyens de pointe, mais pas pour les segments inférieurs. Dans notre modèle, c'est la propriété publique des moyens de pointe qui joue le rôle du prix plafond (entre autres rôles), mais le résultat est similaire puisque les moyens de base disposent toujours d'un pouvoir de marché.

La situation que décrit le modèle apparaît certainement préférable à celle où tous les segments de production sont opérés par des producteurs privés, puisque le niveau de capacité totale limite les risques de délestages et d'effondrement. Cependant, le fait de ne pouvoir atteindre l'optimum de second rang laisse à penser que la situation serait encore meilleure (du point de vue d'un planificateur) si l'ensemble de la production était assurée par un monopole public. Autrement dit, les difficultés à définir des dispositifs réglant les différents problèmes soulevés dans cette partie jettent un doute sur la viabilité de la déréglementation du secteur électrique.

Vers un retour aux monopoles publics ?

L'apparente impossibilité de régler le problème des réserves opérationnelles, impossibilité qui tient la plupart du temps à l'existence d'un pouvoir de marché des producteurs électriques, permet de jeter un regard neuf sur l'organisation du secteur via un monopole public. En effet, les coûts sociaux d'une insuffisance de

capacité sont alors parfaitement intégrés dans les décisions d'investissement, puisqu'ils en sont explicitement les critères de valorisation. Mieux, les différentes contraintes techniques qu'impose le bon fonctionnement des réserves opérationnelles, peuvent être simplement respectées par le biais d'une planification de la construction des centrales, alors que leur traitement par le biais de mécanismes marchand se révèle particulièrement complexe. De plus, le statut de bien public des réserves laisse à penser que la séparation des activités de production de celles de distribution n'est peut être pas si aisée que ce qui a été avancé jusqu'à maintenant. Les moyens de productions éligibles comme réserves constituent de fait un élément fondamental dans la stabilité du réseau, et les facteurs qui conditionnent leur investissement font donc partie à part entière des éléments de valorisation des actifs de distribution. Il y a là matière à de nombreux approfondissements susceptibles de remettre en cause l'organisation des relations entre production et distribution.

Un monopole public supprime de fait tous les risques de manipulation des marchés, dont nous avons montré le rôle fondamental dans la critique des différents dispositifs visant à restaurer les incitations à l'investissement. Il se révèle un mode d'organisation plus simple et plus pratique pour l'exploitation d'un secteur primordial dans le bon fonctionnement du reste de l'économie. Au regard des difficultés à établir des marchés électriques garantissant une bonne allocation des ressources, difficultés soulignées par Joskow et Tirole (2006) dans la conclusion de leur article (« a very challenging task »), un monopole public est simplement une solution qui a fait ses preuves. Cependant, pour pouvoir conclure dans le sens d'un retour au monopole public, il est nécessaire de comparer les coûts de fonctionnement induits par la déréglementation avec ceux d'une gestion publique.

Alors même que la déréglementation du secteur électrique a été poussée par la recherche d'une plus grande efficacité, la contestabilité de la production (par rapport à la distribution) ne semble pas être l'argument décisif pour opter en faveur

de cette organisation. Les caractéristiques intrinsèques du secteur semblent peser de façon plus importante à travers les difficultés touchant le processus d'investissement.

CONCLUSION

L'étude des conditions d'investissement dans le secteur électrique laisse clairement apparaître que celui-ci, sous son mode de fonctionnement actuel, connaît un fort déficit d'incitations à l'investissement. A travers le problème des « revenus manquants » une première reconnaissance, aussi bien théorique qu'institutionnelle, de ce problème est en train d'émerger. Le constat que les marchés électriques ne permettent pas de rémunérer correctement les moyens de pointe (et donc dans une moindre mesure tous les moyens de production) mérite d'être approfondi. Cependant, les premières analyses quantitatives des niveaux de prix sur des marchés proches de l'équilibre (comme nous l'avons réalisé sur le marché français) laissent clairement apparaître un déficit important. Par ailleurs, les causes de ce dernier ne se résument pas uniquement à la mise en place de prix plafonds, puisque d'autres problèmes plus structurels semblent être à la source du problème. Ainsi, les procédures de gestion des déficits de réserves opérationnelles semblent avoir un rôle important dans l'impossibilité d'obtenir des pics de prix couvrant les coûts fixes de nouveaux moyens de pointe. Les marchés électriques, sauf mise en place de mécanismes destinés à traiter ce problème, vont donc certainement connaître de fortes crises dans les années à venir, tant un niveau d'investissement suffisant est central dans leur bon fonctionnement.

Des jalons pour une analyse théorique de ce problème ont été posés dans les pages précédentes. Ils ont permis, à travers la caractérisation du pouvoir de marché par l'investissement ou l'intégration des caractéristiques du fonctionnement des réserves opérationnelles, de disqualifier certains dispositifs visant à remédier au problème des « revenus manquants ». Ainsi, le modèle présenté dans la deuxième partie montre clairement qu'il n'est pas possible de restaurer les incitations à l'investissement en laissant les prix monter jusqu'au coût de défaillance. Plus

généralement, l'existence d'un pouvoir de marché par l'investissement, indissociable de l'absence d'élasticité de la demande et de l'impossibilité de stocker l'électricité, implique qu'on ne pourra régler ce problème par le biais des prix de l'énergie. Les prix plafonds, s'ils permettent de réduire l'exercice d'un pouvoir de marché « traditionnel », n'ont pas ce rôle pour le pouvoir de marché par l'investissement. Il est donc nécessaire d'instaurer une rémunération propre à la capacité si l'on veut pouvoir intégrer les coûts sociaux d'un déficit de capacité dans le processus d'investissement des entreprises.

Par ailleurs, un examen plus approfondi du rôle des réserves opérationnelles laisse apparaître un certain nombre de problèmes supplémentaires. Les réserves opérationnelles ont en effet une fonction primordiale pour le bon fonctionnement des marchés électriques, qui est d'éviter un effondrement du réseau en répondant en temps réel à des variations imprévues de la demande. Indispensables à l'ensemble du réseau, elles sont un bien public. Or, les nouveaux designs de marché de capacité, tout en intégrant un certain nombre d'innovations intéressantes réduisant fortement les possibilités d'exercice d'un pouvoir de marché, ne rémunèrent pas les caractéristiques permettant à un moyen de production d'être éligible comme réserve. D'où le risque que ces designs n'induisent pas (ou pas assez) la construction de moyens éligibles. De plus, il est quasiment impossible d'instaurer des mécanismes de rémunération qui viendraient compléter les paiements de capacité. Du fait de l'hétérogénéité des caractéristiques à prendre en compte, mais aussi de par l'extrême sensibilité des marchés électriques en cas d'insuffisance de réserves. Une solution éventuelle consiste alors à faire construire et opérer les moyens de pointe par le gestionnaire de réseau. Mais cette option, tout en permettant de tenir compte des coûts sociaux d'un déficit de capacité, ne permet pas d'atteindre l'optimum social, les différents segments de production étant interdépendants et laissant toujours la possibilité d'exercice d'un pouvoir de marché.

L'ensemble des problèmes évoqués pointe clairement les difficultés à définir des marchés électriques fonctionnant correctement. Les tentatives pour bâtir des mécanismes permettant de tenir compte des coûts sociaux sont systématiquement contrecarrées par le pouvoir de marché par l'investissement que détiennent de façon quasi « naturelle » les producteurs d'électricité. Même si les analyses précédentes méritent d'être approfondies, le tableau qu'elles dessinent est particulièrement sombre pour l'avenir de la déréglementation du secteur électrique. En effet, son bon fonctionnement est un élément important pour la compétitivité d'une économie. Il constitue également un sujet particulièrement sensible politiquement. On ne fera pas l'économie d'un débat sur l'évolution de ce secteur. Plus tôt il aura lieu, plus il sera possible d'éviter les crises importantes auxquelles semble conduire le fonctionnement actuel des marchés électriques.

Bibliographie

Andersson B., Bergman L. (1995) "Market Structure and the Price of Electricity : an Ex Ante Analysis of Deregulated Swedish Markets." *The Energy Journal*, 16(2): 97-110.

Babusiaux D. (1990) "Décision d'Investissement et Calcul Economique dans l'Entreprise." *Economica* / Editions Technip, coll. « Economie et Statistiques Avancées ».

Boiteux M. (2007) "Les Ambiguïtés de la Concurrence." *Futuribles*, n°331 juin.

Boiteux M. (1956) "Sur la Gestion des Monopoles Publics Astreints à l'Equilibre Budgétaire." *Econometrica*, 24(1), pp. 22-40.

Bolle F. (1992) "Supply Function Equilibria and the Danger of Tacit Collusion: the Case of Spot Markets for Electricity." *Energy Economics*, April , 14(2), pp. 94-102.

Borenstein S. (2002) "The Trouble with Electricity Markets : Understanding California's Restructuring Disaster." *Journal of Economic Perspectives*, vol.16, n°1, pp. 191-211.

Borenstein S. (2000). "Understanding Competitive Pricing and Market Power in Wholesale Electricity Markets." *The Electricity Journal*, vol 13, n°9, july, pp. 49-57.

Borenstein S., Bushnell J. (2000) "Electricity Restructuring : Deregulation or Reregulation ?", *Regulation*, vol. 23, n°2, pp. 46-52.

Borenstein S., Bushnell J. (1999) "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry." *The journal of Industrial Economics*, vol xlvii, No 3 sep, pp. 285-323.

Chaton C. (2001) "Décisions d'investissement et de démantèlement sous incertitude: une application au secteur électrique." *Economie et Prévisions*, 149 (3 jul-sep): 15-28.

Crampes C., Creti A. (2005) "Capacity Competition in Electricity Markets." *Economia delle fonti di energia e dell'ambiente*, n.2, pp 59-83.

Cramton P., Stoft S. (2006) "The Convergence of Market Designs for Adequate Generating Capacity. With Special Attention to the CAISO's Resource Adequacy Problem." manuscript, April, 25, 2006. (<http://stoft.com/p/50.html>)

Cramton P., Stoft S. (2005) "A Capacity Market that Makes Sense." *Electricity Journal*, Vol. 18, Issue 7 pp 43-54. September.

California Public Utilities Commission (2005) "Capacity Markets White Paper." CPUC Energy Division, Sean Gallagher Director, August, 25.

von der Fehr N-H. M., Harbord D. (1993) "Long Term Contracts and Imperfectly Competitive Spot Markets: a Study of the UK Electricity Industry." *The Economic Journal*, 103, pp. 531-546.

Faruqui A., Chao H-P., Niemeyer V., Platt J., Stahlkopf K. (2001) "Analysing California's power crisis." *The Energy Journal*, 22(4): 23-52.

Finon D., Meunier G. (2006) "Market Power in Electricity Markets, Suboptimal Investment and ISO Procurement." GRJM Working Paper 2006-12.

Finon D. (2001) "Integration of European Electricity Market: from Juxtaposition of the National Market to Setting up the Regional Market." *Economies et Sociétés*, série EN n°8, jan-fev.

Glachant J. M. (2000) "Les réformes des industries électriques européennes : à chacun son marché unique." ? *Annales des Mines*, série Réalités Industrielles, août, pp. 10-17.

Glachant J. M. (1998) "Le Système Electrique Britannique : Caractéristiques Quantitatives et Qualitatives du Premier « Marché Ouvert » de l'Union Européenne." ATOM Working Paper 98-02.

Glachant J. M. (1997) "Le Marché Anglais de l'Electricité, Modèle des Echanges Européens ?" *Economie et Sociétés*, série Economie de l'Energie n° 7, 5-6/1997, pp. 153-169.

Green R. (1999) "The Electricity Contract Market in England and Wales." *The Journal of Industrial Economics*, vol xlvii, No 1 mar, pp. 107-124.

Green R. (1996) "Increasing Competition in the British Electricity Spot Market." *Journal of Industrial Economics*, 44, pp. 205-216.

Green R., Newberry D. M. (1996) "Regulation, Public Ownership and Privatization of the English Electricity Industry." in "International Comparisons of Electricity Regulations." Gilbert R. J., Kahn E. P. ed., Cambridge University Press

Green R., Newberry D. M. (1992) "Competition in the British Electricity Spot Market." *Journal of Political Economy*, 100, pp. 929-953.

Hogan, William W. (2005) "On an "Energy-only" Electricity Market Design for Resource Adequacy." Paper prepared for the California ISO, September 23.

Joskow P. L. (2006a) "Markets for Power in the United States : An Interim Assesment." *The Energy Journal*, 27(1): 1-36.

Joskow P. L. (2006b) "Competitive Electricity Markets and Investment in New Generating Capacity." MIT CEEPR Working Paper 06-009, April.

Joskow P. L. (2001) "California's Electricity Market Meltdown." *Economies et Sociétés*, série EN n°8, jan-fev.

Joskow P. L. (1998) "Electricity Sectors in Transition." *The Energy Journal*, 19(2).

Joskow P. L. (1997) "Restructuring, Competition and Regulatory reform in the US Electricity Sector." *Journal of Economic Perspectives*, vol. 11, No 3, sum., pp. 119-138.

Joskow P. L., Tirole J. (2006) "Reliability and Competitive Electricity Markets." *RAND Journal of Economics*, forthcoming.

Klemperer P. D., Meyer M. A. (1989) "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica*, 57, pp. 1243-1277.

Kreps D., Scheinkman J. A. (1983) "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes." *RAND Journal of Economics*, 14, pp. 39-66.

Laffont J-J. (1996) "The French Electricity Industry." in "International Comparisons of Electricity Regulations." Gilbert R. J., Kahn E. P. ed., Cambridge University Press

McDonald R., Siegel D. (1986) "The Value of Waiting to Invest." *Quarterly Journal of Economics*, nov, pp. 707-727.

Newberry D.M. (2000) "Privatization, Restructuring and Regulation of Network Utilities, The Walras-Pareto Lectures.", The MIT, Cambridge, Massachussets.

Pignol C. (1999) "Le Statut de la Concurrence Imparfaite: de l'Amendement des Hypothèses de Concurrence Parfaite à la Représentation d'Agents Price-Makers." Thèse pour obtenir le grade de Docteur en Sciences Economiques de l'Université Paris 1.

Staropoli C. (2001) "Conception de marchés efficaces pour les scteurs déréglementés. Le cas des marchés de gros d'électricité. " Thèse pour obtenir le grade de Docteur en Sciences Economiques de l'Université Paris 1.

Stoft S. (2003) "The Demand for Operating Reserves : Key to Price Spikes and Investment." IEEE.

Stoft S. (2002) "Power System Economics." Wiley IEEE Press.

Tirolle J. (1988) "The Theory of Industrial Organization." Cambridge, MA: MIT Press.

Trebbing H. M. (2001) "New Dimensions of Market Failure in Electricity and Gas Supply." *Journal of Economic Issues*, 35(2).

Varoquaux W. (1996) "Calcul Economique et Electricité." PUF, « Que Sais-je ? », n° 3201

Wolak F.A. (1997) "Market Design and Price Behaviour in Restructured Electricity Markets: an International Comparison", Working Paper, Departement of Economics Stanford University.

Wolak F.A. (2004) "What's wrong with Capacity Markets.", June 23.

Wolfram C. (1999a) "Measuring Duopoly Power in the British Electricity Market." *American Economic Review*, 89 (4 sep): 805-826.

Wolfram C. (1999b) "Electricity Markets: Should the Rest of the World Adopt UK Reforms ?" *Regulation*, Fall, CATO Institute, Washington D.C.

Table des matières

SOMMAIRE	7
INTRODUCTION	8
Partie 1 : L'investissement en moyens de production électrique ...	12
a- Les caractéristiques propres au secteur électrique.....	12
i. Le caractère non stockable de l'électricité.....	12
ii. L'importance des aléas	17
<i>La demande</i>	<i>17</i>
<i>La disponibilité des moyens.....</i>	<i>19</i>
<i>Les conséquences sur l'investissement.....</i>	<i>20</i>
iii. La déréglementation	23
<i>Les différents types de marchés électriques</i>	<i>23</i>
<i>Des dysfonctionnements répétés sur les marchés électriques</i>	<i>28</i>
<i>La pointe de décembre 2001 sur la plaque germanique</i>	<i>28</i>
<i>La crise californienne.....</i>	<i>31</i>
<i>Vers de nouvelles crises ?.....</i>	<i>32</i>
b-Le problème des « revenus manquants » et les incitations à l'investissement.....	35
i. Le problème des « revenus manquants »	36
ii. Mise en évidence du problème sur le marché français	42
iii. Les problèmes de risque	48
<i>Quel est le niveau optimal de capacités ?</i>	<i>50</i>
c-Les modélisations des marchés électriques	54
i. Les modèles de marchés de gros et la problématique du pouvoir de marché.....	56
ii. L'approche en terme de fonctions de profit.....	61
<i>La fonction de profit de court terme et ses propriétés.....</i>	<i>61</i>
<i>Lier gestion des pics de prix et politiques de capacités</i>	<i>65</i>
Partie 2 : Le pouvoir de marché par l'investissement.....	68
a-Choix de modélisation et optimum social	71
i. Le programme d'investissement d'un monopole public.....	75
ii. Analyse du niveau optimal de capacité	79
b-Le duopole déréglementé	81

i. Résolution du programme d'investissement d'une firme	83
ii. Analyse de la solution.....	91
iii. Extension et généralisation du modèle	96
<i>Relâchement des contraintes</i>	96
<i>Généralisation au cas d'un n-oligopole</i>	99
<i>Généralisation à d'autres distributions de probabilité</i>	102
c-Conclusions	103
Partie 3 : Les réserves opérationnelles	107
a-Le rôle fondamental des réserves opérationnelles	109
i. Une politique de gestion des pics de prix est-elle possible ?.....	109
ii. Les réserves opérationnelles sont un bien public	117
b-Les nouveaux designs de marché de capacité	123
i. Les designs « convergents »	123
<i>Le LICAP</i>	127
<i>Le FCM</i>	130
ii. Un modèle de réserves opérationnelles	134
<i>Régimes de fonctionnement</i>	137
<i>Le programme d'investissement optimal</i>	138
c-La propriété publique des moyens de pointe.....	141
<i>Spécifications du modèle</i>	142
<i>L'optimum de Ramsey</i>	143
<i>Le programme d'investissement d'un producteur</i>	144
<i>Vers un retour aux monopoles publics ?</i>	147
CONCLUSION.....	150
Bibliographie.....	153
Table des matières	159